



ESTIMATION OF SOIL ORGANIC CARBON STOCK IN AGRICULTURE AREA
USING SATELLITE BASE REMOTELY SENSED DATA AND MACHINE
LEARNING ALGORITHMS: A CASE STUDY IN THE LOWER PART OF
NORTHEASTERN THAILAND

SUNANTHA OUSAHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE

IN GEOINFORMATICS

FACULTY OF GEOINFORMATICS

BURAPHA UNIVERSITY

2025

COPYRIGHT OF BURAPHA UNIVERSITY

การใช้ข้อมูลการสำรวจระยะไกลและ Machine Learning เพื่อประเมินการกักเก็บคาร์บอนในดิน
ในพื้นที่เกษตร: กรณีศึกษาในภาคตะวันออกเฉียงเหนือตอนล่างของประเทศไทย



สุนันทา อูสาหะ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาภูมิสารสนเทศศาสตร์

คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา

2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

ESTIMATION OF SOIL ORGANIC CARBON STOCK IN AGRICULTURE AREA
USING SATELLITE BASE REMOTELY SENSED DATA AND MACHINE
LEARNING ALGORITHMS: A CASE STUDY IN THE LOWER PART OF
NORTHEASTERN THAILAND



SUNANTHA OUSAHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE
IN GEOINFORMATICS
FACULTY OF GEOINFORMATICS
BURAPHA UNIVERSITY

2025

COPYRIGHT OF BURAPHA UNIVERSITY

The Thesis of Sunantha Ousaha has been approved by the examining committee to be partial fulfillment of the requirements for the Master Degree of Science in Geoinformatics of Burapha University

Advisory Committee

Examining Committee

Principal advisor

.....
(Professor Dr. Zhenfeng Shao)

..... Principal examiner
(Professor Dr. Wolfgang Kainz)

Co-advisor

.....
(Dr. Kitsanai Charoenjit)

..... Member
(Professor Dr. Hong Shu)

..... Member
(Professor Dr. Lanfa Liu)

..... Member
(Professor Dr. Danxia Song)

..... Member
(Professor Dr. Jia Sun)

..... Acting Dean of the Faculty of
Geoinformatics
(Associate Professor Dr. Suchada Pongkittiwi boon)

This Thesis has been approved by Graduate School Burapha University to be partial fulfillment of the requirements for the Master Degree of Science in Geoinformatics of Burapha University

..... Dean of Graduate School
(Associate Professor Dr. Witawat Jangiam)

64910083: MAJOR: GEOINFORMATICS; M.Sc. (GEOINFORMATICS)

KEYWORDS: soil organic carbon stock, remote sensing data, machine learning

SUNANTHA OUSAHA : ESTIMATION OF SOIL ORGANIC CARBON STOCK IN AGRICULTURE AREA USING SATELLITE BASE REMOTELY SENSED DATA AND MACHINE LEARNING ALGORITHMS: A CASE STUDY IN THE LOWER PART OF NORTHEASTERN THAILAND .
ADVISORY COMMITTEE: ZHENFENG SHAO, Ph.D. KITSANAI CHAROENJIT, Ph.D. 2025.

The significance of soil organic carbon (SOC) in mitigating climate change is important. It aids in the removal of carbon from the atmosphere. Soil organic carbon (SOC) affects soil dynamics, a crucial component of soil organic matter (SOM) and soil fertility. SOM enhances soil structure, microbial activity, and plant nourishment. The rapid conversion of biological systems to agroecosystems is stressing the world's soils. Agriculture depletes soil organic matter (SOM) by changing its physical, chemical, and biological qualities. Thus, global scientists want to estimate soil carbon storage. Carbon content was inferred using soil profiles and plant area ratios. Global soil carbon storage estimates intrigue scientists. Soil profile data, spatial interpolation, and soil-landscape model-based DSM predict SOC. Early studies evaluated average carbon density and carbon content using soil profiles and soil-vegetation area ratios. Estimating soil carbon is difficult. Because soil carbon varies by soil type, plant type, and terrain, there are many assessment methods with different results. Tropical landscapes have varied vegetation, soil, and topography. This complicates SOC mapping and characterization across land cover categories and management regimes. This study aims to using remote sensing and machine learning methods to assess soil organic carbon stock in agriculture areas in lower Northeastern Thailand in 2020. We also want to analyze the relationship between environmental variables and remote sensing data in the research locations and compare soil organic carbon estimation machine learning techniques.

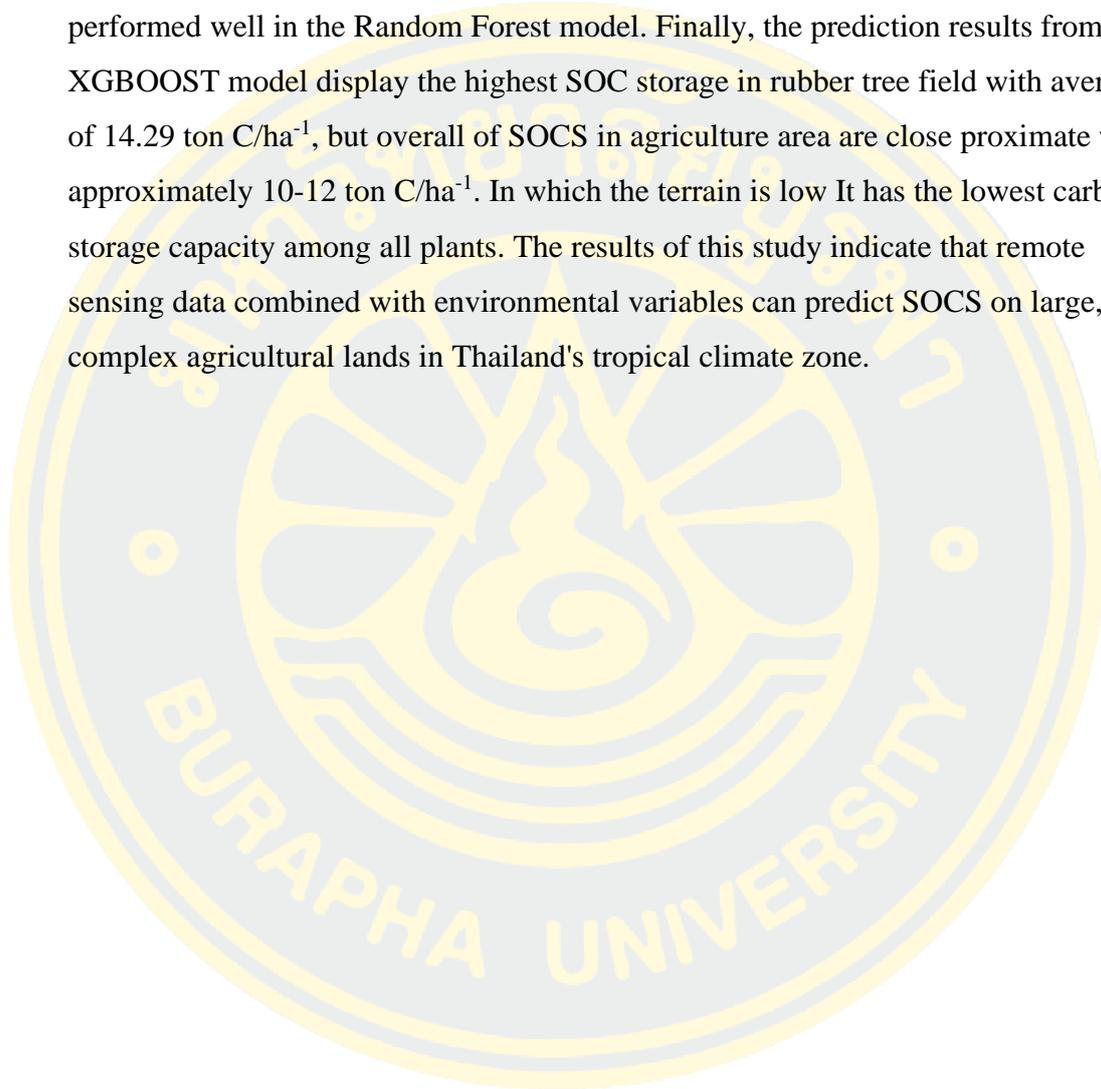
Soil organic carbon storage estimation and monitoring often involve significant fieldwork, soil sampling, and laboratory analysis. Remotely sensed data and machine learning exceed traditional methods. Remote sensing can estimate

regional or global SOC. This helps understand SOC distribution and variations among land cover types and management regimes. In this study, multisource remote sensing datasets, including Sentinel-1 Radar data (VV, VH), Sentinel-2 MSI (Spectral index, band red, and red edge band), SMAP (soil moisture), NASA/CGIAR (topography), NASA/GLDAS-2 (meteorological data), and Land Development Department (LDD) of Thailand (land use map), were used as feature variables in this investigation. Additionally, three machine learning algorithms—RF, SVR, and XGBOOST—must be used to estimate soil organic carbon, not only to improve the accuracy of SOC estimates but also to process large amounts of data rapidly, thereby enhancing the efficiency of SOC estimation. This can be beneficial for estimating SOC over large regions or for monitoring frequently over time. The most essential component of data is the soil sampling dataset; 234 soil samples collected via ground survey and laboratory method from LDD in 2020 with difference land use type and soil series were evaluated using a machine learning model.

Multiple sources of remote sensing data were gathered using the open-source Google Earth Engine Python API software which is a highly effective platform for geospatial analysis and remote surveillance. All remote sensing data were analyzed utilizing cloud-based processing, mean value calculation, soil and vegetation index extraction, etc., and resampling at a resolution of 30 meters. Similarly, soil sampling datasets were utilized to calculate the total soil organic carbon stock (ton C/ha⁻¹) in the study area based on soil series data using the kriging method. XGBOOST and RF models processed the model using the cuML GPU package, which provides speed, scalability, accuracy, and flexibility, while SVR used TensorFlow and permutation important to calculate features important in each model. The model was validated by dividing the training data into 70% and the testing data into 30%. The performance of the model was then assessed using R², RMSE, and MAE.

The result demonstrates that slope, vegetation index, rainfall, soil moisture, and land use play a significant role in SOCS prediction. According to the results of machine learning, XGBOOST has the highest SOCS prediction accuracy among the three models, with R² = 0.803, RMSE = 5.569, and MAE = 3.204, in

combination with the top five features: EVI, SLOPE, Rainfall, TVI, and Soil moisture. The significance of features of SOCS prediction is exceptional in three models: SLOP, EVI, Land use, and Soil Moisture variable. However, RI, NDVI, and Temperatures performed effectively in the SVR model, whereas Red-edge2 performed well in the Random Forest model. Finally, the prediction results from XGBOOST model display the highest SOC storage in rubber tree field with average of 14.29 ton C/ha⁻¹, but overall of SOCS in agriculture area are close proximate with approximately 10-12 ton C/ha⁻¹. In which the terrain is low It has the lowest carbon storage capacity among all plants. The results of this study indicate that remote sensing data combined with environmental variables can predict SOCS on large, complex agricultural lands in Thailand's tropical climate zone.



ACKNOWLEDGEMENTS

During my master's degree studies, I am indebted to my supervisors, Prof. Zhenfeng Shao, for their invaluable counsel, continuous encouragement, and forbearance. His guidance and expertise have been instrumental in enhancing my understanding of this complex dissertation. Their vast expertise and vast experience have inspired me throughout my academic research and daily life. I would like to thank Soil Resources Survey and Research Division, Land Development Department, Land development department in Thailand for providing soil sampling data for machine learning model training. I would like to express my sincere gratitude to Mr. Zeeshan Afzal for his exceptional support and invaluable knowledge in the field of machine learning algorithms. I am truly grateful for his unwavering assistance, which has significantly contributed to my growth and development in the field. Thank you, Mr. Zeeshan Afzal, for your dedication and for sharing your expertise. I would like to thank all four SCGI master's degree cohort members. Their assistance and support have made my time studying and living in China so enjoyable. Faculty of Geoinformatics, Burapha University, Thailand, and The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Photogrammetry and Remote Sensing College, Wuhan University, People's Republic of China, is to be thanked for providing me with the means to pursue my education. In addition, I would like to thank the researchers of Thailand's Geo-Informatics and Space Technology Development Agency (GISTDA) for their invaluable assistance in shaping my experimental methodologies and evaluating my results. Lastly, I'd like to express my appreciation to my parents. It would have been impossible for me to conclude my studies without their extraordinary support and understanding over the past few years.

Sunantha Ousaha

TABLE OF CONTENTS

	Page
ABSTRACT	D
ACKNOWLEDGEMENTS	G
TABLE OF CONTENTS	H
LIST OF TABLES.....	J
LIST OF FIGURES	K
CHAPTER 1 INTRODUCTION.....	1
Statements and significance of the problems.....	1
Research questions.....	3
Research objectives	3
Contribution to knowledge	3
Scope of research work.....	4
Thesis structure	5
CHAPTER 2 LITERLATURE REVIEW	9
The carbon cycle and soil carbon sequestration in agroecosystem	9
Factors influencing the organic carbon storage of soils	13
Measurements of soil organic carbon storage.....	15
The status of soil organic carbon storage	17
Physical characteristics of the study area	20
Spectral reflectance and spectral signature of soil.....	24
Satellite base remotely sensed data.....	25
Machine learning algorithms in soil carbon stock.....	28
A review of research on using remote sensing for soil carbon stock	31
CHAPTER 3 RESEARCH METHODOLOGY	34
General background of study area	34
Workflow of research	35

Soil sampling and laboratory analysis	37
Statistical data analysis of soil organic carbon	41
Acquisition and processing of data.....	41
Predictive model	48
Accuracy assessment	51
Soil organic carbon spatial distribution and validation	51
CHAPTER 4 RESULT	53
Statistical descriptions	53
The importance of remote sensing data and environment variables.....	57
Performances of machine learning models.....	59
Soil organic carbon spatial distributions and validation.....	63
CHAPTER 5 CONCLUSION AND FUTURE WORK	72
Conclusion	72
Future work and suggestions	73
REFERENCES	74
BIOGRAPHY	79

LIST OF TABLES

	Page
Table 1 Land use in the lower part of Northeastern Thailand in 2020	22
Table 2 Soil texture characteristics in the study area separated by landform	23
Table 3 Band and spectral index related estimation of soil organic carbon storage...27	27
Table 4 List of remote sensing and environment variables	42
Table 5 Sentinel-1A data properties for filtering.....43	43
Table 6 Sensor spectral and spatial characteristics of the red and red-edge band	44
Table 7 Spectral index information for Sentinel-2A data.....45	45
Table 8 Land use classification in the study area	47
Table 9 Three predictive models selected in this study.....48	48
Table 10 The levels of SOC stock in this study.....52	52
Table 11 Descriptive statistics for the entire SOC stock dataset.....54	54
Table 12 Descriptive statistical analysis of remote sensing and environmental variables	55
Table 13 The hyperparameters of the three models with the highest performance	60
Table 14 Comparison and evaluation of predicted model performance.....60	60
Table 15 Soil organic carbon stock in five agriculture land from XGBOOST model64	64
Table 16 SOCS spatial distribution in difference level of paddy field.....66	66
Table 17 SOCS spatial distribution in difference level of casava	67
Table 18 SOCS spatial distribution in difference level of sugarcane	68
Table 19 SOCS spatial distribution in difference level of rubber.....69	69
Table 20 SOCS spatial distribution in difference level of perennial crop	71

LIST OF FIGURES

	Page
Figure 1 Overview of the thesis structure.	6
Figure 2 The slow carbon cycle from the atmosphere to the lithosphere	10
Figure 3 The carbon cycles via land and atmosphere through the photosynthesis	11
Figure 4 The soil organic matter stabilization; Green arrows indicate inputs, blue microbial systems, and white physicochemical processes.....	12
Figure 5 Soil carbon sequestration in soil profile	14
Figure 6 The procedure of collecting soil samples in the field.....	16
Figure 7 GSOCmap (v1.5.0) represents soil organic carbon globally	18
Figure 8 Soil organic carbon (SOC) in the upper 30 centimeters.....	19
Figure 9 The study area (a) Thailand, (b) Northeast part, and (c) the study area	20
Figure 10 Spectral reflectance of various amounts of organic carbon in wetland soils. (Jianli Ding, 2018)	24
Figure 11 Backscattering of different wavelengths against vegetation and soil in difference bands of SAR; X band, C band, and L band.....	26
Figure 12 Diagram of random forest classifier	29
Figure 13 Maximum-margin hyperplane and margins for an SVM trained with samples from two classes.....	30
Figure 14 Quadratic functions to convert lower-dimensionality space to a higher-dimension space in Kernels Support Vector Machine	31
Figure 15 Location of the study area on the overview map of Thailand	35
Figure 16 The overall workflow of this study	36
Figure 17 Soil sampling point in the study area (234 soil samplings).....	37
Figure 18 Example of soil pit and soil profile (a) casava field (b) soil profile using auger tools (c) collecting soil sample in the bags (d) characteristics of soil texture in rice fields.....	38
Figure 19 Example of soil collection by undisturbed core method	38

Figure 20 A photograph of a location with numerous gravel and fragments of rock. (a) gravel in surface soil (b) remove some huge rock fragments (c) fine and moderately plant root in surface soil, and (d) collect soil and gravel	40
Figure 21 The skewness histogram of soil organic carbon data	53
Figure 22 The histogram of remote sensing and environmental variables data	56
Figure 23 Heatmap of feature collection and soil organic carbon storage using Pearson coefficient analysis.....	57
Figure 24 Permutation importance of each models (a) XGBoost model; (b) RF model; and (c) SVR model.....	58
Figure 25 SOCS at 0–30 cm predicted from the XGBoost model.....	61
Figure 26 SOCS at 0–30 cm predicted from the Random Forest model	62
Figure 27 SOCS at 0–30 cm predicted from SVR model.....	62
Figure 28 Land use type in agricultural area in the study area	63
Figure 29 SOCS spatial distribution of paddy field.....	65
Figure 30 SOCS spatial distribution of casava	66
Figure 31 SOCS spatial distribution of sugarcane.....	68
Figure 32 SOCS spatial distribution of rubber tree	69
Figure 33 SOCS spatial distribution of perennial crop.....	70

CHAPTER 1

INTRODUCTION

Statements and significance of the problems

Global warming is a significant issue that causes negative effects such as rising sea levels, warming oceans, and the alteration of animal populations and habitats. Carbon dioxide levels in the atmosphere are at their highest in millions of years due to the use of fossil fuels, which humans are returning to the atmosphere in only a few hundred years. Although fossil fuel emissions declined during the pandemic, atmospheric carbon dioxide still reached record high levels (IPCC, 2022). Global warming is expected to speed up the breakdown of organic matter, leading to the release of more carbon dioxide into the atmosphere. Land use change and fossil fuel consumption are the two major factors that disrupt the carbon cycle (IPCC, 2019). Changes in land use can alter the carbon balance, especially when a controlled system replaces a natural one. Active SOC is a sensitive indicator of SOC changes after land use changes. Changing land use has a negative impact on carbon stock. Sustainable land use and management practices can act as a significant carbon sink in agricultural areas.

Remote sensing technology can estimate soil organic carbon (SOC) storage, which is crucial for soil health, carbon sequestration, and ecosystem services. The use of satellite or airborne sensors can provide information on vegetation cover, soil moisture, and land use patterns to estimate SOC content (Angelopoulou et al., 2019). Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) can measure the amount of green vegetation, which is related to SOC storage (Bilgili, 2020). LULC data can also estimate SOC storage by mapping different land use types such as cropland, forests, and grasslands. Overall, remote sensing provides a cost-effective and efficient method for monitoring and managing soil health and carbon sequestration efforts. Nevertheless, estimating soil organic carbon (SOC) storage in agricultural areas of tropical climates using remote sensing and machine learning faces several challenges. These include the high spatial heterogeneity of SOC, limited availability of ground-truth data, frequent cloud cover, and rapid changes in SOC

levels due to multiple cropping cycles and different land use practices (Pandey, 2017). To overcome these challenges, machine learning models must be developed to capture these dynamics and account for temporal changes in SOC storage.

Northeastern Thailand is a region with diverse landscapes and agriculture practices, including rice cultivation, forestry, and livestock rearing. However, the region faces several challenges related to soil and soil degradation, including soil erosion and pollution from the overuse of chemical fertilizers and pesticides. Additionally, low soil organic matter content and soil fertility pose significant problems for agricultural production and climate change mitigation. Addressing these issues requires integrated soil management practices such as organic farming, conservation tillage, crop rotation, and cover cropping. By improving soil health and carbon storage, these practices can promote sustainable agricultural production in the region. Therefore, there is an opportunity to employ remote sensing and machine learning techniques to estimate and monitor land use changes and associated soil organic carbon changes more efficiently. These methods can enable rapid and accurate estimation and monitoring of soil organic carbon levels, which is critical for understanding soil health and promoting sustainable land management practices.

The aim of this study is to estimate soil organic carbon stock in agricultural areas of Thailand, using remote sensing data suitable for the tropical climate areas. The study utilizes important data such as Sentinel-1 C-band Ground Range Detected (GRD) scenes and Sentinel-2 MSI to create vegetation and soil indices. Additionally, environmental factors such as Digital Terrain Elevation Data (DTED) level 2 products, meteorological data, land use, and soil moisture were incorporated as input data for the predictive model. The research employed three machine learning algorithms, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR), to estimate the soil organic carbon stock. Ultimately, the study will calculate the soil organic carbon storage across various land use types in agricultural areas of the lower part of Northeastern Thailand.

Research questions

1. How much soil carbon storage that estimated using remote sensing data and machine learning in the study area
2. What is the most relationship between environment variable and remote sensing data
3. How much accuracy of using satellite image can estimate soil organic carbon in agriculture areas
4. What is the best machine learning algorithm for estimation of soil organic carbon storage

Research objectives

To study soil organic carbon stock in agriculture areas by using remotely sensed data and machine learning method under climate and land use change including relationship between remote sensing data, topographic, and meteorological data to produce SOC stock database for supporting appropriate soil management and land use planning for planner, decision maker, agricultural development in Agriculture areas, Thailand.

1. To estimate soil organic carbon stock by using satellite base remotely sensed data and machine learning algorithms in agriculture area.
2. To examine the correlation between feature importance and soil organic carbon stocks using various machine learning techniques.
3. To compare different machine learning algorithms for estimating soil organic carbon stock in agriculture areas in Thailand.

Contribution to knowledge

1. The opportunity to develop innovation and method of soil organic carbon stock mapping and estimation which can monitor rapidly changing of SOC in agriculture soil of Thailand.
2. Acknowledgment of the SOC stock and understanding of carbon behavior in agroecosystems under climate and land use change can be used as SOC database to

develop sustainable practices for increasing of advantageous plant nutrients and soil carbon stock.

3. The chance to produce CO₂ emission-reduction projects can be traded and sold, and used by industrialized countries to earn certified emission reduction (CER) credits under the Kyoto Protocol.

Scope of research work

1. Scope of study area

Soil organic carbon store in agricultural regions in Chaiyaphum, Khon Kaen, Maha Sarakham, Buri Ram, and Nakhon Ratchasima Provinces (a total of around 59,780 sq. km.) is all that this research is concerned with estimating. The geological characteristics of the region include a central plain surrounded by rocky hills to the west and south. Geology is situated inside the Khorat basin. The majority of soil is poor in fertility due to long-term leaching, and the majority of soil is generated by residuum of the Korat group, which consists of many quartzes, depending on the parent material, produces almost coarse to coarse-textured soil, sandy loam, and well-drainage. The province has the greatest number of Roi-et soil series, Korat soil series, and Satuk soil series, respectively. In addition, the province is spearheaded by two saline soil series—the Kula Ronghai and the Udon. Most research areas are agricultural, including rice, cassava, sugarcane, maize, etc. Average temperatures in the area vary from 32 to 22 degrees Celsius, making the climate a Tropical Savannah.

2. Science data scope

The spectral index data was obtained from Sentinel-1 with VV, and VH polarizations and Sentinel-2A satellite with the addition of B2, B3, B4, B5, B6, B7, B8 B8A, B10, B11, and B12 (accessed on 01 January - 30 May 2020). The Sentinel-2 MSI L2 GRD product data was selected to calculate vegetation and soil index. The digital terrain elevation data (DTED) level 2 products with a resolution of 30 m were used to extract various terrain attributes. From this DEM data, the terrain variables generated were as follows: elevation and slope. Moreover, soil moisture data, topography, and climate data were acquired and processed using Google Earth Engine

through the Google Earth Engine Python API on Google Colab. Land use map at scale 1:25,000, level 3 in agriculture areas from Land Development Department. Finally, soil pit samples were dug, and undisturbed soil core samples were taken by a cylindrical core sample. We collected 234 soil samples of each land use type and soil texture for training set and testing set data.

Thesis structure

The goal of this study aims to estimate soil organic carbon stock for agricultural areas in the lower part of Northeastern in Thailand. There are officially main five chapters in this thesis. Firstly, chapter 1 presents the summary of the estimation of soil organic carbon stock in Thailand, the significance of problems of the study, the research objectives, contribution of knowledge, scope of research work, and conceptual framework. Secondly, chapter 2 illustrated with a literature review in the carbon cycle and soil carbon sequestration in agricultural ecosystem including soil organic carbon stock measurements, factors affecting soil organic carbon stock, and satellite base remotely sensed data and machine learning algorithms. Thirdly, chapter 3 shows the data in use, methodology, and the process of this research. Fourthly, chapter 4 displays the analysis result of soil organic carbon stock from field, the soil organic carbon stock using remotely sensed data and environment data including correlation from the former and latter data, and comparison machine learning. Finally, chapter 5 is a conclusion and recommendation for future research in the dissertation.

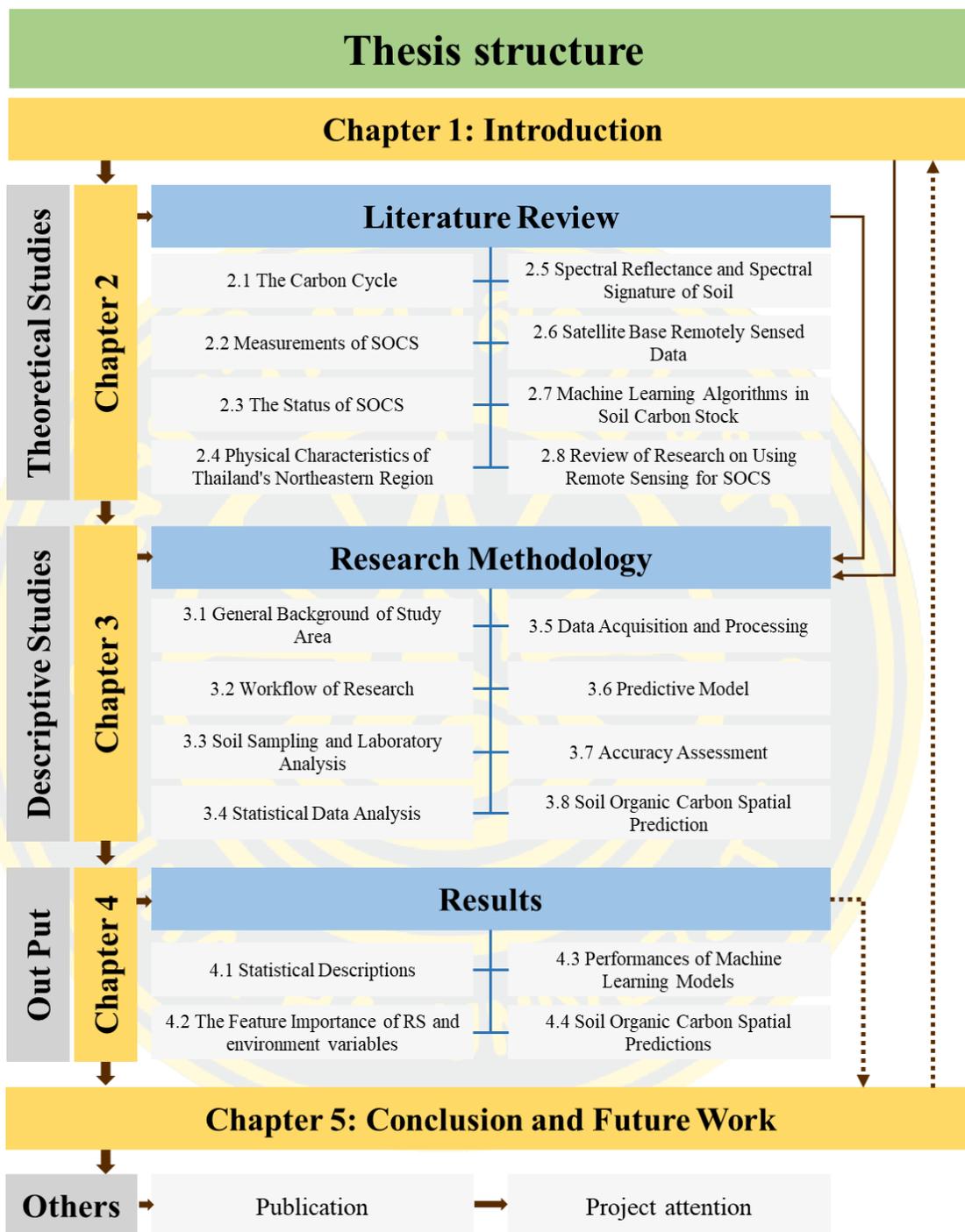


Figure 1 Overview of the thesis structure.

CHAPTER 2

LITERLATURE REVIEW

In the first part of my literature review, carbon cycle was discussed as well as the relationship that exists between the carbon cycle and soil organic carbon sequestration, including the factors influencing the organic carbon storage (sequestration) of soils (Topics 2.1). This will aid researchers achieve a better understanding of how carbon is stored in the soil. Secondly, the principle of measuring soil organic carbon storage will describe the theory and formula for the soil organic carbon stock. (Topic 2.2). The Status of Soil Organic Carbon Storage on a Global Scale will be presented in this section of topic 2.3, followed by the Physical Characteristics of Thailand's Northeastern Region, and study area, including location, climate, geography, topography, land use, and soil resources in section 2.4. In sections 2.5, 2.6, and 2.7, respectively, the principles of spectral reflectance, remote sensing data, and machine learning model will be discussed. In conclusion, we will examine a study of Research on Remote Sensing for Soil Carbon Stock from the past to the present (Topic 2.7).

The carbon cycle and soil carbon sequestration in agroecosystem

The carbon cycle on Earth allows for the continuation of life. The vast majority of living things give off carbon dioxide (CO₂) through a process called respiration. Carbonate rocks make up the lithosphere and can be found in coal, natural gas, and petroleum. Minerals, peat, biological detritus, and muck are all parts of the pedosphere's composition. Trees act as a repository for photosynthesis. Aquatic ecosystems are nourished by the shells and marine plants that are found in the hydrosphere of lakes and oceans. The carbon cycle is kept under check by carbon pools. Carbon molecules are constantly being traded between the biosphere, geosphere, pedosphere, hydrosphere, and atmosphere. As a result of chemical weathering, potassium, calcium, magnesium, and calcium or sodium ions are carried by rivers to the ocean, where they combine with carbon dioxide to form blue carbon and calcium carbonate. (FAO, 2021) Calcium carbonate comes from plankton and

shell builders. Limestone traps carbon from the organism's breakdown. Mud layers enclose organic carbon. Pressure and heat compress silt and carbon to form shale over millions of years. Shale is sometimes replaced with organic carbon from oil, coal, or natural gas. Volcanoes release carbon (NSSO, 2018).

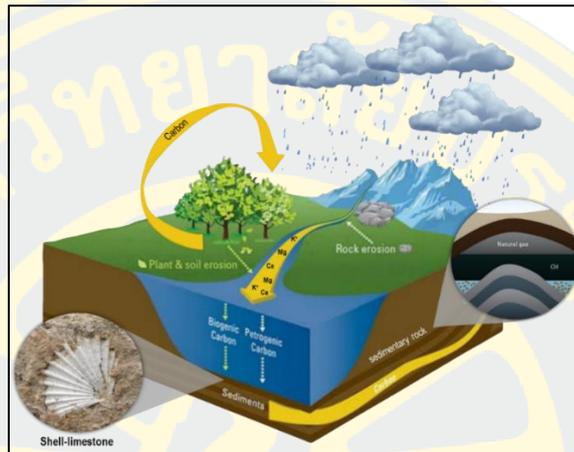


Figure 2 The slow carbon cycle from the atmosphere to the lithosphere

Earth life accelerates carbon cycles. Plants and phytoplankton gather carbon dioxide from the atmosphere and use solar energy to convert CO_2 and water into sugar (CH_2O) and oxygen in the rapid carbon cycle. Cellular respiration releases plant carbon. Sugar fuels plants. Planting sugar fuels creatures, including humans. Post-growing season bacteria devour plants and plankton. Plants burn. Oxygen and sugar each release water, CO_2 , and energy. Reactions release CO_2 . Atmospheric carbon dioxide fluctuations can foretell plant growth season due to the rapid carbon cycle. Fast and slow carbon cycles maintain atmospheric, terrestrial, and water carbon levels. Reservoir carbon changes influence climate-affected reservoirs.

Soils are the greatest carbon reservoir on Earth, with approximately 60% of Carbon existing as SOM and 40% as inorganic carbon. Only the oceans, with 40,000 billion tons contain more carbon than soil organic matter at 1,500 billion tons. Soil contains more carbon as organic and inorganic carbon than the atmosphere and plants combined, approximately 760 billion tons and 560 billion tons, respectively. In temperate and arctic climates, the amount of SOC and CO_2 absorption can be

controlled by soil carbon sequestration. Higher temperatures in an atmosphere of sufficient water accelerate the decomposition of soil organic matter, reduce carbon storage in the apathetic and passive pools, and increase the loss of carbon through photosynthesis (K. Nabiollahi, 2019). In temperate climates, there is typically less organic carbon in the soil than in frigid climates. In semi-arid and arid climates, however, carbon sequestration in the soil is possible through the transmutation of atmospheric CO_2 into organic compounds such as secondary carbonates; however, the rate of formation of inorganic carbon is relatively slow. As released carbon diffuses into the atmosphere as carbon dioxide, agriculture is a significant contributor to rising greenhouse gas emissions.

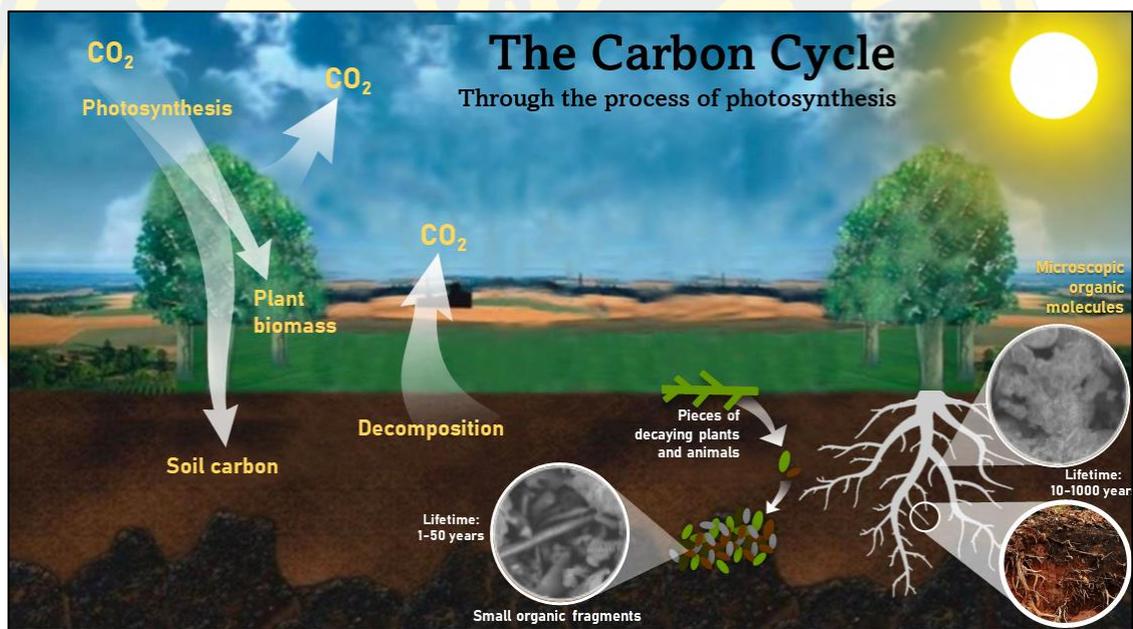


Figure 3 The carbon cycles via land and atmosphere through the photosynthesis

It is possible for agricultural soils, particularly agricultural soils that have been managed, to absorb the emissions of greenhouse gases and contribute to the reduction of these emissions. Through the implementation of best farming practices, agriculture not only helps to the goals of water quality and soil conservation, but it also contributes to the increase of the quantity of organic carbon that is stored in the soil and to the mitigation of the impacts of the emission of greenhouse gases on

climate change. The goal of global efforts has been to reduce releases of greenhouse gases while simultaneously increasing the amount of carbon that can be stored (sequestered) away from the atmosphere. This is accomplished by converting part of the carbon that is produced during photosynthesis in plants and storing it in soil pools.

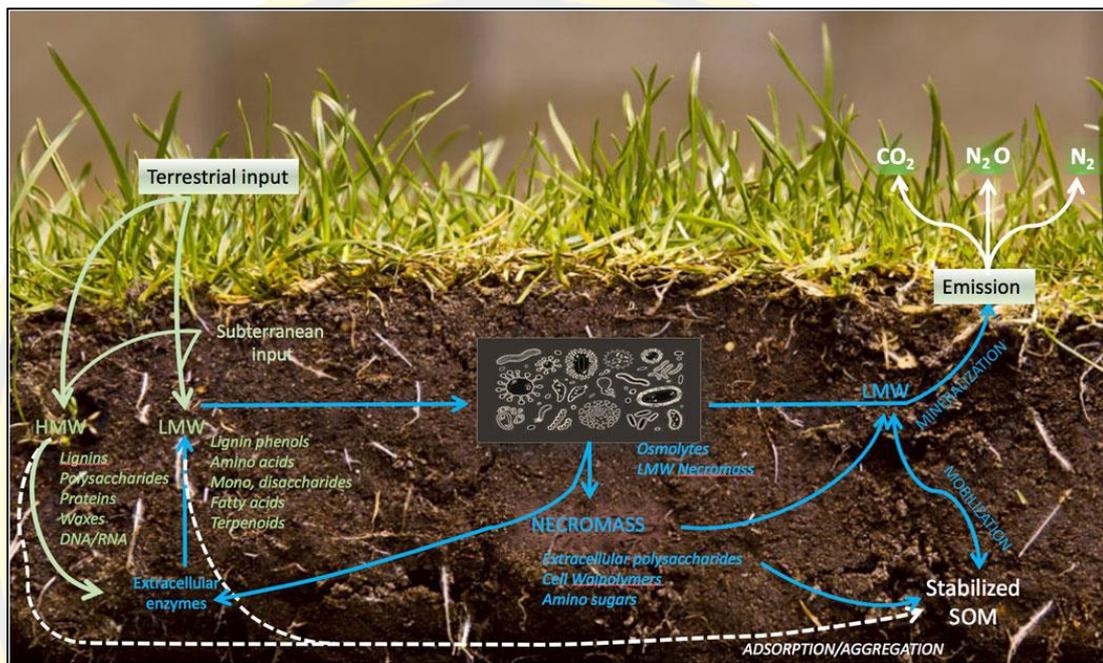


Figure 4 The soil organic matter stabilization; Green arrows indicate inputs, blue microbial systems, and white physicochemical processes.

SOC is sequestered by complex biological, chemical, and physical processes. The populations of soil microorganisms are influenced by debris and root exudates from both above- and below-ground locations. Some LMW compounds are directly ingested by soil microorganisms, which then produce extracellular enzymes to degrade HMW plant inputs. LMW can be rapidly depleted by the process of respiration mineral interaction, or microorganisms mobilizing stable SOM molecules (Ajami M, 2016). Most of the soil organic matter may be produced by neuromas. Land use changes reduce global soil SOC reservoirs. The harvesting and burning of residue decrease organic material components, cultivation increases breakdown

stages, and soil erosion reduces subsoil and organic matter, thereby decreasing SOC stocks.

Factors influencing the organic carbon storage of soils

Agricultural management practices also influence the SOC rate of turnover and equilibrium levels. Typically, agricultural residue retention minimizes soil erosion and water loss throughout fallow periods. Nonetheless, crop remains left on the ground's surface of the land have negligible long-term effects on soil carbon stocks. Carbon is lost predominantly through the mineralization of organic matter and soil erosion; the majority of SOC is stored in the eroding black soil agricultural landscape at the foot of the slope. (L. A. Jones, 2016). One of the primary factors contributing to the system's consistency is controlled by a number of different parts, all of which engage in intricate communication with one another. These variables include the qualities of the soil that lies beneath the surface, the current environmental conditions (with regard to both temperatures and rainfall), as well as how land is utilized and management strategies.

1. Soil Properties

In general, the SOC concentration decreases as the depth of the soil profile increases (P annop, 2016). At depths of more than 50 cm, the concentrations of SOC are found to be higher, shown as Figure 6. The level of soil organic carbon (SOC) can vary from one soil type to another due to differences in temperature, soil mineral composition, topography, and management strategies. The texture of the soil has an effect not only on the total quantity of SOC that is stored in the soil but also on the amount of net mineralization of SOM. The decomposition of organic matter occurs more quickly in sandy soils compared to clay soils. The pH of the soil is the most important factor. The hydrogen ion activity of the soil, which is measured by pH, is what controls how effective soil microbial enzymes are. When the pH of the soil is at about 6.7, organic matter mineralization happens at the quickest rate. In addition, soil factors such as bulk density, pores, and pore size all influence the availability of water and oxygen in the soil, which in turn influences the breakdown of SOC. The percentage of the soil's pore space that is filled with water should be

between 50 and 75% for optimal microbial respiration. The amount of water that is contained in the soil's pores is governed by the amount of water in the soil, but it is also strongly impacted by the amount of clay and organic matter in the soil.

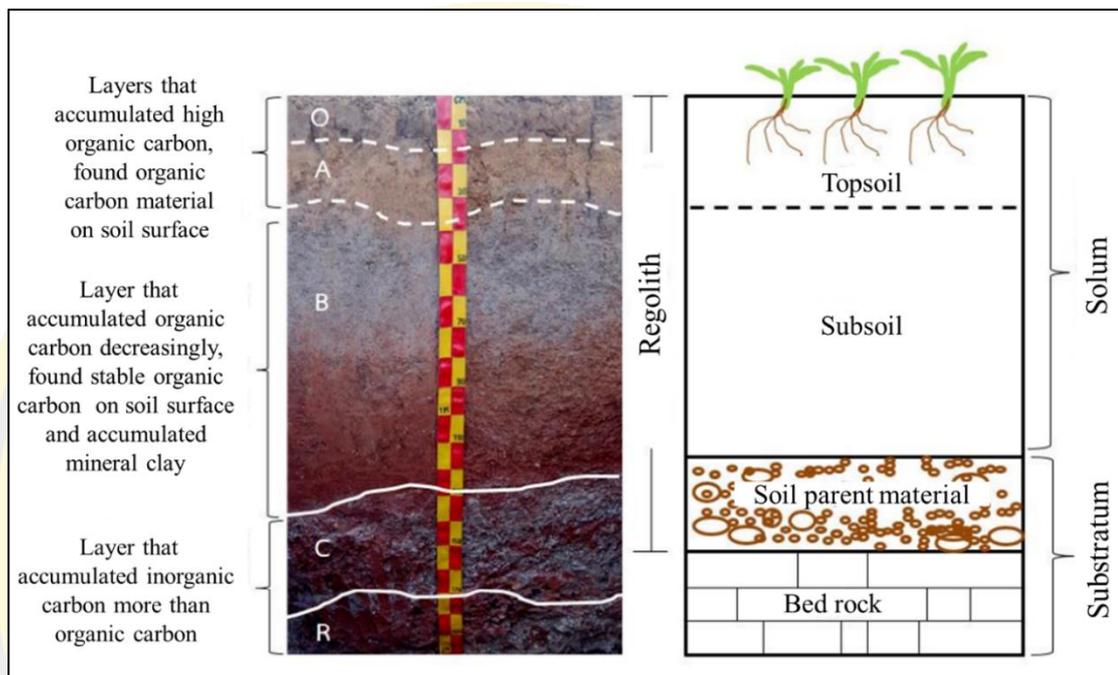


Figure 5 Soil carbon sequestration in soil profile

2. Climate

Temperature and humidity affect soil organic matter inputs and losses. Warmer climates have less organic soil C. Temperature and moisture alter SOC respiration, affecting soil C storage. Temperature directly affects SOC decomposition. Tropical regions decompose SOC faster. Soil carbon loss was significant when soil organic carbon decreased mean annual temperature (MAT). (Hartley et al., 2021) Soil moisture affects the flow of water, gas and solute diffusion, and microbiological activity. Precipitation affects SOC stocks through vegetation. Considerable potential sources of carbon under a warmer and wetter environment and SOC loss were mostly regulated by future climate change and root-zone SWC. SOC loss's hump-shaped reactions to warming temperatures and SWC change showed that the SWC could mitigate warming's impacts on SOC loss. (Fubo Zhao, 2021)

3. Land management

As soon as cultivation begins, the vast majority of undeveloped soils observe a precipitous decline in SOC levels. This is primarily the result of decreased carbon inputs into the soil structure as well as an increase in the breakdown of plant wastes and SOM in agricultural systems. The amount of organic carbon that is stored in the soil can be affected by any management strategies that either directly or indirectly impact the activity of microorganisms and the rate at which they decompose, as well as those that promote production and the return of plant wastes to the soil. Management activities such as tilling increase the rate of decomposition of soil organic carbon because they make the organic matter more available to the microorganisms in the soil that are responsible for its breakdown. When compared to tilling, low or no disturbance with residue preservation has a stronger capacity to either keep carbon in the soil or reintroduce it. (Minnikova et al., 2022) In most cases, the quantities of organic carbon that are stored in the soil can be raised by management practices such as fertilization, irrigation, and the usage of recycled organic resources. This type of increase is possible because organic carbon is a very stable form of carbon.

Measurements of soil organic carbon storage

Carbon in the soil cannot be measured. Some methods are more straightforward and require fewer presumptions and blunders than others. Careful, repetitive sampling in the field followed by analysis in the laboratory using the dry combustion method, also referred to as elemental analyses, is the highest standard for soil carbon assessment. Dry combustion or elemental analysis is the most precise and cost-effective soil carbon assay. The majority of research indicates that soil carbon changes most swiftly in the organic matter fraction, so if you detect a change, it is likely in the organic carbon fraction. If carbon dioxide constitutes a significant proportion, knowing the proportion of organic and inorganic carbon in the soil will help you detect the change. Dry combustion and Walkley-Black are the most common techniques for the precise storage of soil organic carbon.

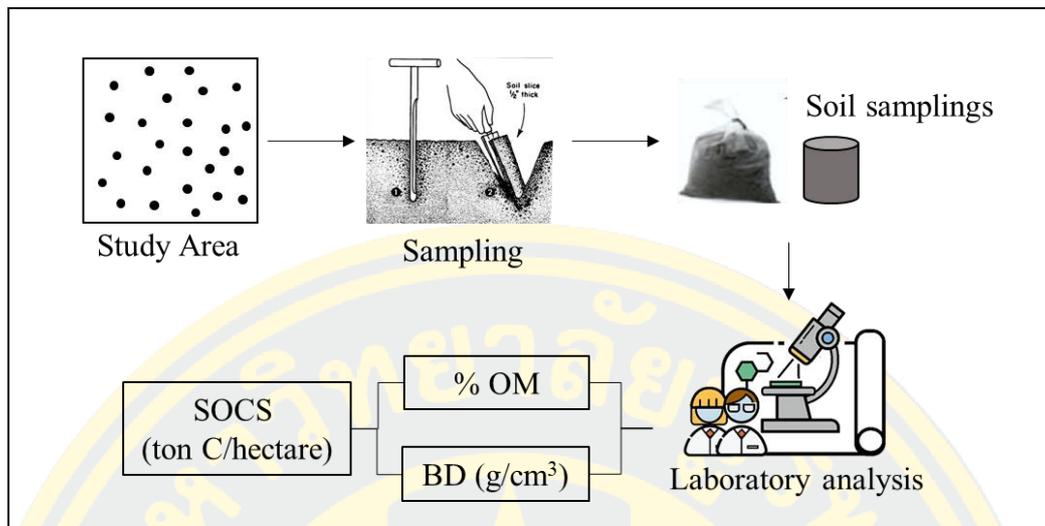


Figure 6 The procedure of collecting soil samples in the field

1. Dry combustion or elemental analysis

The most accurate laboratory soil carbon test is dry combustion using an elemental analyzer from Leco, Perkins-Elmer, Elementary, or Carlo Erba. Autoanalyzer's use dry combustion to analyze TC. Autoanalyzer's enhance precision and adaptability. Autoanalyzer's analyze carbon, nitrogen, and sulfur.(FAO, 2019) These instruments heat a small sample of dry, pulverized soil to a temperature that is around 900 degrees Celsius, at which time they measure the quantity of carbon dioxide gas that is created as a consequence of the combustion of the soil. The results are shown as a percentage of the overall quantity of carbon that was discovered in the sample. The dry combustion test involves oxidizing the organic matter, charcoal, and carbonates present in the soil to determine the total amount of soil carbon.

2. Walkley-black

The Walkley-Black (WB) method is both speedy and minimally demanding of resources in its equipment needs, making it an attractive alternative to other methods of combustion, whether they involve wet or dry combustion. (Litong Chen, 2015). When two volumes of highly concentrated H_2SO_4 and one volume of a

solution of $K_2Cr_2O_7$ are mixed together, the heat of dilution makes it easier for oxidizable organic matter on the ground to be oxidized by chromic acid during the presence of sulphuric acid. This occurs when the heat of dilution is combined with the presence of sulphuric acid. The quantity of dichromate reduced can be used as a starting point for calculating the amount of material that was oxidized. The loss on ignition (LOI) and Walkley-Black tests that have traditionally been used are less exact. The term "loss on ignition" refers to the weight loss experienced by a dehydrated soil sample while it was heated in a combustion chamber or muffle furnace at temperatures ranging from 360 to 450 degrees Celsius for several hours. The Walkley-Black procedure is a type of wet chemical analysis that makes use of potassium dichromate. The entire carbon content is not measured by these tests. In most cases, the Walkley-Black test does not offer a comprehensive analysis of charcoal and may fail to account for some types of organic materials. Both are unable to quantify inorganic carbon. The interest in soil carbon is a relatively new phenomenon when seen from the point of view of either the functioning of the biosphere or the effects of climate change. For the goal of determining the appropriate dosage of herbicide to apply, numerous laboratories are experienced in conducting analyses of the organic matter in the soil. In this context, the presence of organic matter in the soil is undesirable since it lessens the efficiency of chemicals on living plants; a loss on ignites or Walkley-Black test is generally utilized.

The status of soil organic carbon storage

The historically rapid increase in worldwide agricultural production, which is continuing strong now, has had a significant impact on the cycles of carbon, water, and nutrients. Both the conversion of land for agricultural purposes and the production of food through agriculture have been and continue to be significant contributors to the rise in atmospheric levels of carbon dioxide (CO_2). Nearly half of the world's potential vegetation land surface has been turned into agricultural land, pastures, and rangelands. This conversion took place on a global scale. Estimates of global soil carbon stocks, trends, and sequestration potential are now central to important estimations that are currently ongoing in various international fora. These estimations include estimates from the Food and agriculture organization of FAO,

NASA (SMAP L4 Carbon Product), and UNFCCC. These estimations are particularly relevant in the context of a warming climate and land use change.

Food and agriculture organization of the united nation (FAO, 2022) in 2022, GSOCmap is the first global soil organic carbon map created through a consultative and participatory approach including member countries. Member nations under the International Technical Panel on Soils and Global Soil Partnership Administration created the map. 2020 soil organic carbon stocks (ton C ha⁻¹) are estimated. Figure 7 compares the average annual sequestration rate (ton C ha⁻¹ yr⁻¹) for SSM1, SSM2, and SSM3 scenarios with 5%, 10%, and 20% C input. The 4p1000 Initiative: Soils for Food Security and Climate estimates that croplands could absorb 26–53% of their yearly emissions, or 0.90–1.85 Pg C/yr. (Zomer et al., 2017) The significance of regions that have seen intensive agricultural production, such as those in North America, Europe, and India, as well as those in Africa, such as Ethiopia, is underlined here. Soil carbon storage and the protection of current soil carbon stores are key mitigation techniques for attaining the global target of 2 degrees Celsius established by the Paris Agreement on Climate Change. Soil carbon sequestration offers a variety of benefits, one of which is an increase in the food supply.

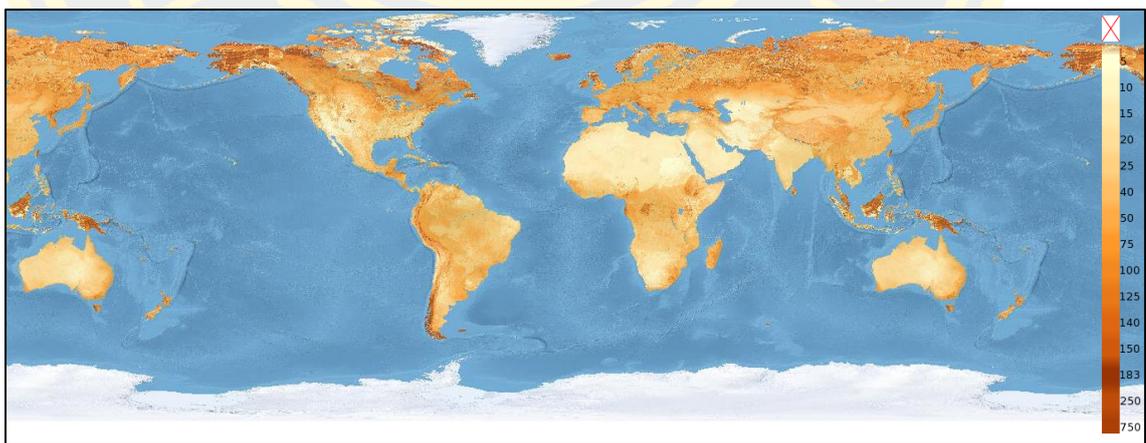


Figure 7 GSOCmap (v1.5.0) represents soil organic carbon globally

In 2021, José Padarian et al. (José Padarian, 2021) looked into the possibility of adding more storage space on top of the existing croplands around the

world. According to the study, the increased storage of sulfur dioxide potential between 29 to 67 Pg C in the subsoil of worldwide croplands is comparable to only 2 to 5 years of emissions balancing and accounts for 32% of agricultural calculated previous carbon debt of 92 Pg caused by the destruction of natural ecosystems. This information was gleaned via a scientific investigation. Since SOC is affected by temperature, it is anticipated that this potential will decline by 18% by the year 2040 because of climate change.

The UNFCCC used ESRI ArcGIS software and data from the SoilGrids250 database to map soil organic carbon (SOC) distribution, aiding efforts to mitigate greenhouse gas emissions. SOC distribution varies with latitude due to differences in sunlight and photosynthesis. Most global SOC is stored in northern permafrost and boreal regions. Over 20 years, SOC in the top 30 cm of soil increased by 0.27–0.54% under medium and high scenarios, equating to 0.012–0.027% annually. Globally, agriculture contributes over 140 million Pg C to the top 30 cm of soil, about 10% of the world's SOC pool.

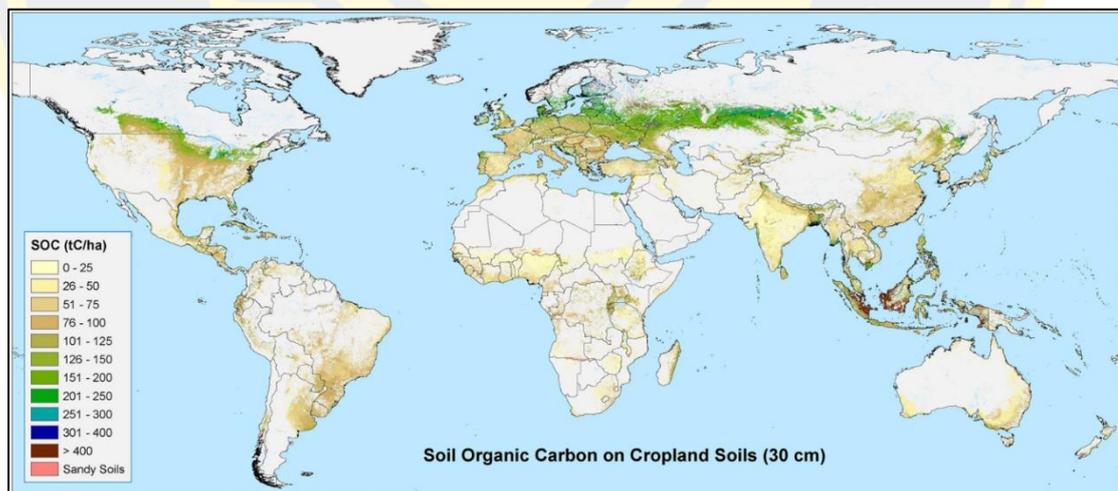


Figure 8 Soil organic carbon (SOC) in the upper 30 centimeters

Figure 8 shows that North America, Eurasia, and Europe store the most carbon on farmland, each having over 21 Pg C and contributing over 50% of the world's SOC inventories. Central America, North Africa, and the Australian region

contain 4.6% of the world's SOC, 6.48 Pg C. Western, South, Southeast, and East Asia have trace amounts of 4.38 to 9.14 Pg C, less than 2% of the global total. Despite agricultural productivity, South America has a mild 9.42 Pg C. Africa has roughly 12 Pg C, 8.5% of the globe overall, with the highest level in the Eastern and Central areas. Russia preserves 21.9 Pg C of SOC in agriculture, 17% of the global total, which is followed by the US (18.9), China (8.4), India (6.4), and Brazil (5.0).

Physical characteristics of the study area

1. Location

The study area, located in the lower part of Northeastern Thailand (Isan), spans approximately 59,780 square kilometers and comprises five provinces. It is Thailand's largest and most populous region, bordered by Laos to the north and east, Cambodia to the south, and separated from Central and Northern Thailand by the Phetchabun Mountain range to the west. The Mekong River forms the international boundary with Laos and Cambodia. Geographically, the region lies between 14°–19°N latitude and 101°–106°E longitude, encompassing the Khorat Basin in the southwest and the Sakon Nakhon Basin in the northeast, divided by the Phu Phan Mountain Range.

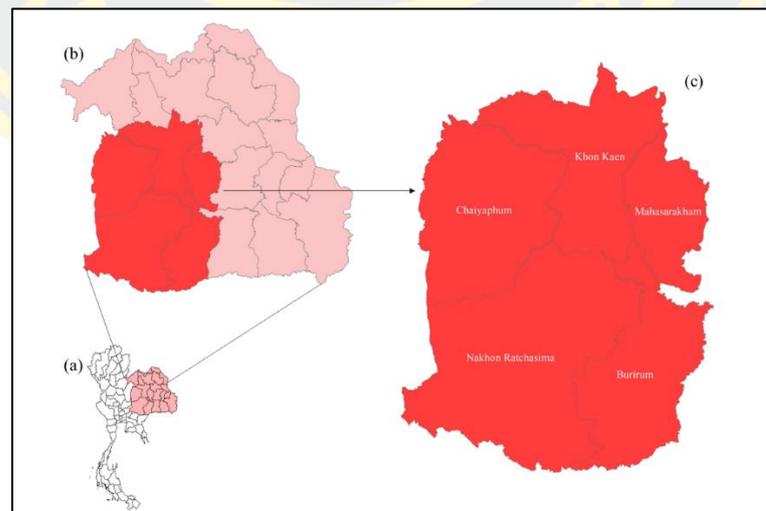


Figure 9 The study area (a) Thailand, (b) Northeast part, and (c) the study area

2. Climate and temperature

The lower part of Northeast region has tropical low rainfall and tropical savannah climate (Aw). Northeast monsoon winds from the Sunda shelf bring cold, dry weather from November to February. The Southwest monsoon winds from the Indian Ocean normally arrive in June-September, but they might start early in May. Most affected years vary in intensity and duration. Average temperature is 32.6°C to 22.2°C, relative humidity 72.8%, rainfall 1,412.4 mm, and 119 rain days. Thus, three seasons.

2.1 Rainy

The wet weather began in late May or early June and continued into October. Despite the fact that the southwest monsoon has an effect on the northeastern region, according to the Petchabun range, which is located to the west of the region, and the Phanom Dongrak range, which is located to the south and acts as a large barrier with slightly affected rainfall. Depression Storm, which was located over the Gulf of Tonkin, is the source of most of the rain that has fallen in this region. It had just made landfall in upper Vietnam and was heading in the direction of northeastern Thailand.

2.2 Winter

The first month of winter was October, and it lasted all the way through February. The winds from China are blowing through the countryside of Thailand. The North and the Northeast experienced a decrease in temperature because of the introduction of a dry and chilly wind that blew across the Northeast.

2.3 Hot season

The beginning of the hot season was in February and lasted until late May. Long distances between land and the gulf of Thailand contribute to progressively high temperatures and extremely dry weather conditions. These conditions are caused by the passage of monsoon winds from the southwest that move through the South China Sea and the Gulf of Thailand.

3. Geography and topography

Northeast is situated on the Khorat Plateau, which includes the Sakon Nakhon basin and the Khorat basin and has been inhabited since prehistoric times. The

average elevation range is between 100 and 200 meters, and there are five main mountain ranges, including the Petchabun range, Dong Phrayayen range, Sankamphaeng range, Phanom Dongrak range, and Phu Pphan range. (Sukanlaya Choenkwan, 2014)

The region's geology is dominated by the Mesozoic Khorat Group, comprising six formations (Nam Phong to Khok Kruat) deposited by meandering streams and floodplains. The group is overlain by the Maha Sarakham Formation and Late Cretaceous Phu Tok sandstone. Older deposits include Late Triassic Hui Hin Lat fluvial-lacustrine sediments and Silurian-Tertiary igneous rocks of the Loei-Phetchabun Fold Belt. The Khok Kruat Formation dates to the Late Early Cretaceous (Aptian).

4. Land use

According to surveying data and land use mapping (LDD, 2020), there are five primary categories (Table 1). The Land Development Department's two-year-ahead land use projection for 2020 served as a benchmark for soil management practices. Agriculture is the predominant land use there, accounting for 4,218,476.49 hectares.

Table 1 Land use in the lower part of Northeastern Thailand in 2020

Type	Areas	
	Hectares	Percentage
Agriculture	4,218,476.49	70.57
Forest	860,669.03	14.40
Urban	414,319.43	6.93
Miscellaneous areas	276,217.85	4.62
Water	208,317.21	3.48
Total	5,978,000.01	100

5. Soil resources

In 2015, the Land Development Department (Satira Udomsri, 2015) reported that soil resources are associated with four categories of geological characteristics: flood plain, alluvial terrace, peneplain, and rocks-controlled landform. There are 178 important soil series at a scale of 1:25,000 in this region. Spodosols, Oxisols, Vertisols, Ulsisols, Mollisols, Alfisols, Inceptisols, and Entisols are the eight soil orders. As shown in Table 2, soils texture from a variety of upland and lowland regions in the region have different textures. The majority of soils have a sandy loam texture that is frequently low and dependent on soil organic carbon (SOC) levels.

Table 2 Soil texture characteristics in the study area separated by landform

Mapping units	Areas	
	Hectare (ha ⁻¹)	Percentage (%)
Sandy Loam Soil	2,684,997.91	44.91
Silty Loam Soil	159,087.46	2.66
Silty Clay Loam Soil	457,752.82	7.66
Silty Clay Soil	133,442.21	2.23
Sandy Clay Loam Soil	48,818.97	0.82
Loamy Sand Soil	664,305.23	11.11
Loam Soil	279,535.72	4.68
Gravelly Sandy Loam Soil	5,966.30	0.10
Gravelly Clay Loam Soil	99,691.41	1.67
Clay Loam Soil	501,319.32	8.39
Clay Soil	34,804.31	0.58
Slope Complex	743,992.46	1.32
Urban	18,096.46	12.45
Miscellaneous	61,325.81	0.30
Water	78,958.77	1.03
Others	5,904.85	0.10
Total	5,978,000.00	100.00

The majority of the soil has low fertility due to leaching that took place for an extended period of time, and the majority of the soil was created by residuum of the Korat group, which is composed of numerous quartzes (SiO₂). This, combined with the fact that the parent material generates nearly coarse to coarse textured soil, sandy loam, and well drainage, makes the majority of the soil nearly coarse to coarse in texture.

Spectral reflectance and spectral signature of soil

Most of the flux that strikes the surface of soil is either reflected or absorbed, while just a little portion is transmitted. There is a positive association between reflectance and wavelengths, and most soils have reflective qualities that are comparable to one another. The moisture content, organic content, texture of the soil, structure of the soil, and iron oxide content of the soil are the five features of soil that affect its reflectance qualities. These characteristics are listed in order of importance. All these characteristics are connected to one another in some way. For instance, the texture of the soil, which refers to the percentage of sand, silt, and clay particles, is connected to both the structure of the soil, which refers to the organization of these particles into aggregates, and the capacity of the soil to retain moisture.

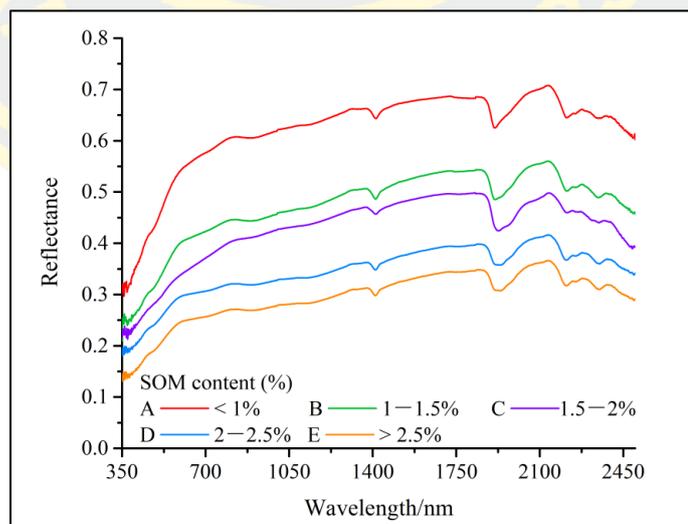


Figure 10 Spectral reflectance of various amounts of organic carbon in wetland soils. (Jianli Ding, 2018)

In the case of soil organic matter, in the soil is opaque, the presence of this dark component lowers the reflectance of the soil, but only up to a point where it contains approximately 4–5 percent organic matter. When the proportion of organic matter in the soil is higher than 5 percent, the soil becomes dark, and any subsequent additions of organic matter have no impact on the soil's reflectance.

Satellite base remotely sensed data

The use of remotely sensed data has made it possible to develop one of the most effective devices for surveying and mapping the Earth at many spatial scales, including local, regional, and global. There are a great number of important applications that are currently dependent on data and images obtained from satellites able to anticipate and estimate agricultural works such as crop sown area estimates, crop disease identification, soil characteristics, soil moisture, soil erosion, flood impact, and other such things in addition to estimating and monitoring changes in the environment and the weather. The use of satellite imagery in agriculture helps cover a large area of land and can be helpful in determining the current state of crops. Band and spectral index are the two primary elements derived from remotely sensed data that are associated to substantial soil organic carbon estimation.

1. Synthetic aperture radar data

SAR (Synthetic Aperture Radar) technology has the advantage of both daylight and nighttime all-weather imaging. The European Space Agency's (ESA) newly released Sentinel satellites feature brief revisit time, high spatial resolution, free download, and high-quality mechanical validation data. They can provide vast quantities of free optical and SAR remote sensing data with a high spatial resolution for prediction. The property that can capture the soil–vegetation relationship to predict soil chemical properties.(Demir, 2021) Multitemporal Sentinel-1 images display fluctuating correlations between backscattering coefficients and SOC.

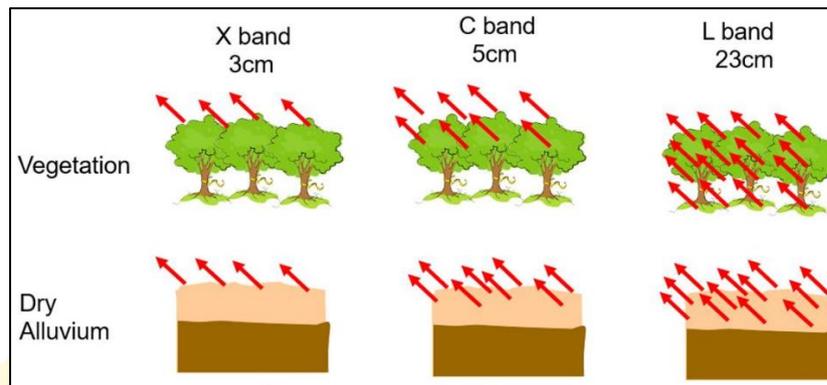


Figure 11 Backscattering of different wavelengths against vegetation and soil in difference bands of SAR; X band, C band, and L band

The utilization of the polarization approach with VV, HH, VH, and HA polarization is an essential component of radar imaging and is its most distinguishing feature. The radar backscatter is determined not only by the parameters of the sensor, such as the incidence angle, polarization, and wavelength, but also by the parameters of the target, such as the surface roughness, the structure of the vegetation, and the dielectric qualities. (Figure 12) Moreover, polarization can occur in either the sending or receiving of signals due to the fact that objects can reflect signals in a certain direction. As a result, the strength of the backscattering can be exploited as a method for the collection of information concerning the surface attributes. Data gathered by radar satellites can be obtained at a number of scales, ranging from meters to kilometers, because of the utilization of the synthetic aperture radar (SAR) principle. This allows for the data to be used for a wide range of applications. These days, the C-band is utilized for the purpose of predicting the organic carbon content of the soil. The application of SAR data at the L-band for predicting soil organic carbon (SOC) is currently restricted and very seldom described in scientific literature. This is primarily because of the complexity, diversity, and availability of the data.

The relationship between VV and VH backscattering, and soil organic carbon (SOC) is complex and influenced by factors such as soil type, moisture content, surface roughness, and vegetation cover. VV backscattering is sensitive to soil roughness and moisture, while VH is sensitive to vegetation cover and subsurface

soil structure. Higher SOC content results in lower backscattering values for both polarizations, but other factors such as soil texture and mineralogy can also affect the signal. The relationship may vary depending on the radar signal frequency used. It is important to consider these factors and use additional methods to accurately estimate SOC in the soil.

2. Optical satellite data

SOC shows the strongest relationship with reflected electromagnetic radiation in the visible (around 600 nm), NIR, and SWIR regions. Satellite imagery, such as Sentinel-2 MSI and Landsat-8 OLI, has demonstrated strong potential for SOC modeling, prediction, and mapping. Incorporating variables like vegetation, topography, climate, and geology further enhances spectral models. Techniques such as physical, statistical, geostatistical, and machine learning approaches combine these data for SOC monitoring. Indices like NDVI and MSAVI are effective for biomass estimation and strongly correlate with SOC prediction.

Table 3 Band and spectral index related estimation of soil organic carbon storage.

	Band/Spectral Index	References
Band	B2, B4	
Vegetation Index	The Normalized Burn Ratio (NBR2)	(Klara Dvorakova, 2020)
	Normalized Difference Vegetation Index (NDVI)	(Yangchengsi Zhang, 2019) , (Azamat SULEYMANOV, 2021)
	Transformed Vegetation Index (TVI)	(Azamat SULEYMANOV, 2021)
	Difference Vegetation Index (DVI)	(Xiaohang Li, 2021)
	Redness Index (RI)	(Bayartungalag Batsaikhan, 2020)
	Color Index (CI)	
	Brightness Index (BI)	
	Enhanced Vegetation Index (EVI)	
Soil Index	Modified Soil Adjusted Vegetation Index (MSAVI)	(Shuai Wang, 2021)
	Soil Adjusted Vegetation Index (SAVI)	

Machine learning algorithms in soil carbon stock

Machine learning models are an important tool in agricultural system and agroecosystem assessment, increasingly used by researchers, environmentalists, academician, food producers, and soil surveyor. Examples of their applications include captured parameters and quality of soil, seed quality, fertilizer application, plant growth, genetic and environmental conditions, irrigation and harvesting. In agriculture, machine learning has the potential to improve not just water and plant nutrient management but also environmental protection and resistance to soil deterioration. Machine learning algorithm methods help decision-making systems in Agriculture and make it possible to predict the status of soil, plant, and environment condition including costs incurred.

1. Random forest (RF) model

Random forest is a famous Supervised Machine Learning Algorithm for Classification and Regression issues. Bagging is the ensemble technique utilized by random forest to generate a distinct training subset from sample training data with replacement, and the final output is determined by the majority vote. Random forest is a collection of decision trees; still, there are a lot of differences in their behavior such as comparatively slower, created from subsets of data and the final output is based on average, and randomly selects observations. Typically, the random forest algorithm consists of four steps: (1) In Random Forest, n random records are selected from a data set containing k records. (2) Separate decision trees are created for each sample. (3) Every decision tree will produce an outcome. For classification and regression, respectively, the final output is determined by majority voting or averaging.

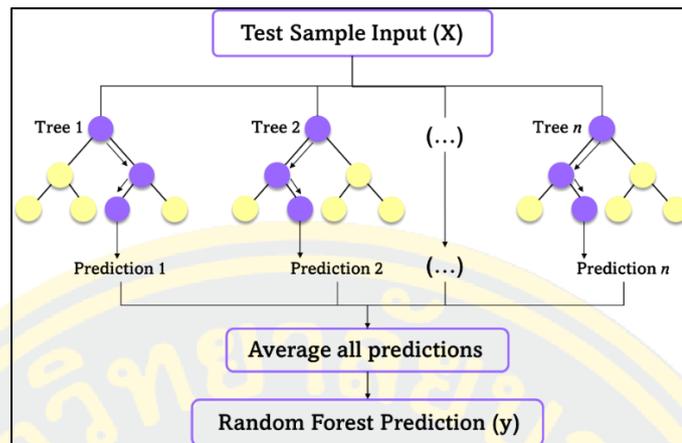


Figure 12 Diagram of random forest classifier

Random forests use hyperparameters to either improve the performance and predictive ability of models or to speed up the model. Observing hyperparameters improves predictive accuracy: 1) $n_estimators$ are number of trees the algorithm builds before averaging the predictions. 2) $max_features$ are maximum number of features random forest considers splitting a node. 3) $mini_sample_leaf$ is determining the minimum number of leaves required to split an internal node.

The following hyperparameters enhance speed: n_jobs specifies how many processors the engine is permitted to use. If the value is 1, only one processor can be used, but if it is -1, there is no *limit.random_state*: controls the sample's randomness. If the model has a fixed value of the random state, the same hyperparameters, and the same training data, it will always produce the same *outcomes.OOB* is an acronym for out of the bag. It involves random forest cross-validation. This one-third of the sample is not used to train the data, but rather to assess its efficacy. These samples have been extracted from sack samples.

2. Extreme gradient boosting (XGBoost) model

XGBoost is an open-source package implementing gradient boosting, typically using decision trees as base learners in an ensemble approach. It combines outputs from multiple models to enhance prediction accuracy. Decision trees, fundamental to XGBoost, operate by sequentially splitting data using binary decisions

(e.g., yes/no) to predict outcomes at their leaves. Splits are chosen to maximize the separation of target values, iterating until a stopping criterion, such as maximum tree depth, is met. This process ensures efficient and accurate classification or regression.

3. Support vector regression (SVR)

However, the support vector machine is mainly applied for classification problems; it is one of the most popular supervised learning algorithms applied to the field of machine learning for regression problems. Support vector machine is another basic algorithm that every expert in machine learning should possess because it achieves high precision with less computing capacity. The objective of the SVM algorithm is to generate the most appropriate line or decision boundary that divides n -dimensional space into classes to ensure future information can be readily classified. This best decision boundary is referred to as a hyperplane, following as Figure 15.

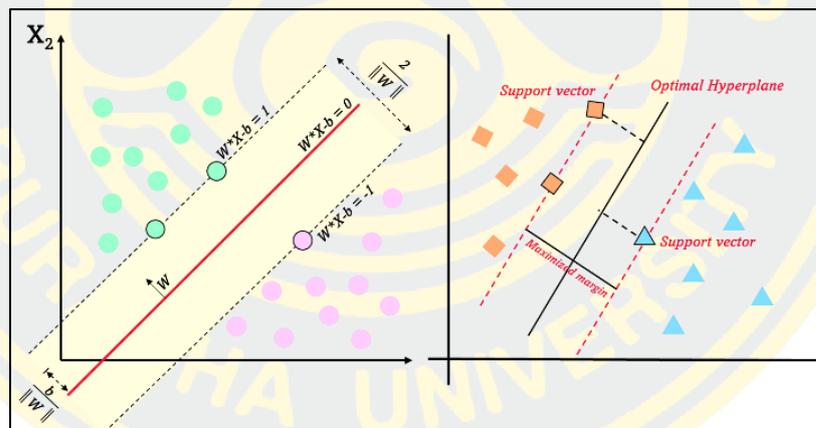


Figure 13 Maximum-margin hyperplane and margins for an SVM trained with samples from two classes.

There are two types of SVM such as Linear SVM and Non-Linear SVM. Only in the case where the data can be completely linearly separated the option of using Linear SVM, whenever data points have been perfectly linearly separable, it indicates that they can be divided into two groups using just one straight line as the dividing line. On the other hand, If the data cannot be separated into two classes using a straight line (if the data is in two dimensions), afterward we use non-linear SVM to

categorize it. This means that when the data points cannot be separated into two classes using a straight line, then we use some advanced techniques such as kernel tricks. Because linearly separable data points are unusual in real-world applications, we resort to the kernel trick to discover solutions to these problems(Saini, 2021).

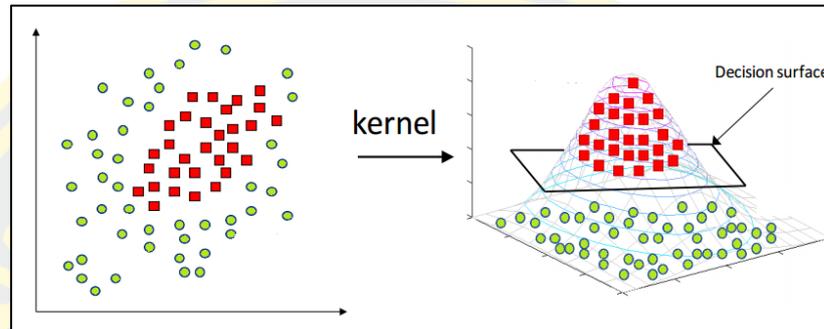


Figure 14 Quadratic functions to convert lower-dimensionality space to a higher-dimension space in Kernels Support Vector Machine

Kernels in Support Vector Machine is the most interesting feature of SVM that it can even work with a non-linear dataset and for this, “Kernel Trick” makes it easier to classifies the points. According to Figure 16, a hyperplane is incapable of producing images that correctly classify points. Using some quadratic functions to convert this lower-dimensionality space to a higher-dimension space will allow us to locate a decision boundary that clearly separates the data points. Kernels and hyperparameter optimization functions can accomplish these problems.

A review of research on using remote sensing for soil carbon stock

Soil organic carbon (SOC) holds a key part in the Carbon-cycle as one of the important carbon pools on Earth. As a result, carbon sequestration may be able to mitigate climate change. As a component of Organic Matter (OM), SOC influences the physical, chemical, and biological properties of a soil ecosystem, as well as its structure and water and nutrient retention. Scientists are becoming interested in SOC. Since the early 2000s, SOC publications have grown. SOC monitoring is time-consuming and expensive, thus researchers studied alternate ways that may be used in

varied settings. Recent efforts have focused on determining whether various Remote Sensing (RS) methods and machine learning techniques can be used to estimate various soil properties, including SOC, in a quick, inexpensive, and non-destructive manner.

The review examines studying of soil organic carbon stock related to remote sensing data, modeling, algorithm, including machine learning refers to on remotely sensed satellite data in passive sensors (optical satellite images) and active sensors (Radar, SAR, Lidar). All the research on estimating, mapping, monitoring has taken place in the past and most of it in just the last few years of new technology and methods.

Since the 1998 Kyoto Protocol, soil carbon sequestration and its impact on C fluxes have received more attention. Article 3.4 discusses methods to establish baselines of C stocks and modalities, rules, and guidelines for adding and subtracting human-induced activities related to changes in greenhouse gas emissions by sources and removals by sinks in agriculture soils, land use, and forestry categories from the assigned amounts of C sequestration. here is in addition to forests and land use change (Article 3.3), which the COP may adopt, and is particularly relevant to soil C because here is the only area other practices can be introduced (UNFCCC, 1998). An experimental field for organic carbon assessment is an appropriately traditional method and useful for a long time. Laboratory tests are the gold standard remains careful, repeated field sampling followed by laboratory analysis by the dry combustion method, often called elemental analysis (Donovan, 2013). Organic carbon content was determined by Walkley and Black method using collecting soil samples. Soil samples were collected from a pit of 30cm by 100 cm pit and undisturbed soil core samples by a cylindrical core sampler (P. Ghimire, 2018). On the other hand, the role of subsurface soil horizons as predictors of the SOC stock in the whole soil profile was analyzed. The observed average value of SOC stocks in the studied subsurface horizons reaches 73% of the whole soil (Manuel Rodríguez-Rastrero, 2022). Even though the traditional methods are still useful and high accuracy, soil sampling should be carried out at a time and in a stabilized ground, not recently ploughed. Moreover, the ground water table should be preferably below the soil depth to be sampled. In croplands, the sampling must be carried out at least four weeks after

the last ploughing at a time when the upper soil layers are firm and wet and allow for using a soil auger or automated corer. In addition, periods following cultivation operations (tillage, disking, seeding, fertilizers application etc.) must be avoided (Dominique Arrouays, 2018).

At the beginning, Remote sensing has played a vital role in this endeavor during the last five decades by quantifying carbon fluxes and stocks. Remote sensing data acquired in a broad wavelength range (visible, infrared, and microwave) of the electromagnetic spectrum have been used to estimate carbon fluxes and/or stocks. Optical satellite multispectral imagery, Spaceborne and Airborne, started to be used extensively in quantitative SOC characterization. For example, an integrated satellite–airborne mapping approach that supports high-resolution carbon stock assessment and monitoring in tropical forest regions (Asner, 2009). Moreover, it can also provide the data to segment a site according to its soil heterogeneity, while extending existing datasets of soil properties to support digital soil mapping (V.L. Mulder, 2010). The primary attribute can be predicted with kriging within strata, or some combination of regression analysis and kriging or cokriging.

Within the past 10 years from 2012 to 2022, remotely sensed data and machine learning has been recognized as a promising technology in estimation soil carbon storage. Nonlinear models may be more effective at elucidating the complex soil-environment relationship than linear models. Prediction accuracy reduces from Unmanned Aerial Systems (UASs) to satellite platforms, though advances in machine learning techniques could further assist in the generation of better calibration models (Theodora Angelopoulou, 2019). Landsat 8 ETM⁺ and Random Forest (RF) models were used to predict SOCS in Iran that proficient estimated in the topsoil and subsurface in croplands and forestland (K. Nabiollahi, 2019). Similarly, Boqiang Xie (2022), SOC estimation was carried out using Sentinel-2A outperformed under The best results were obtained by the XGBoost algorithm. $R^2 = 0.804$, $RMSE = 1.771$, $RPIQ = 2.687$. On the other hand, BRT models is the best predictions for SOC content with $R^2 = 0.470$ using Sentinel-2, Sentinel-3 and Landsat-8 images (Zhou et al., 2021). Sources of RS data and machine learning algorithms accuracy are differentiated in terms of their areas, spectral, model parameters, and temporal resolution that consecutively specifies their accuracy and the field of application.

CHAPTER 3

RESEARCH METHODOLOGY

As it is indicated in the title, this chapter explains various methodologies that were used in gathering data and analysis which are relevant to the research. The methodologies will include areas such as the location of the study, general background of the lower part of Northeastern, Thailand, workflow of research, data acquisition, type of data, preprocessing of satellite imagery, environment data of soil organic carbon distribution, machine learning methods and its management. In addition, validation and generating soil organic carbon map is mentioned in this chapter also.

General background of study area

The study area comprises Chaiya Phum, Khon Kaen, Maha Sarakham, Buri Ram, and Nakhon Ratchasima Province and is located in the northern part of northeastern Thailand between approximately 101°18'E to 103°51'E longitudes and 14°12'N and 17°08'N latitudes, covering an area 59,780 square kilometers. Geographical characteristics of the region include a central plain, the rugged Phetchabun range to the west, and the Phonomdongrak Range to the south. And PhuPan Range is situated in the northern portion of the region. Alluvial plains exist along the Mun and Chi rivers. The Mahasarakam Formation is the primary source of salt among the subterranean minerals, which are predominantly composed of sandstone, shale, limestone, and siltstone. According to soil progenitor materials, this region's soil resources consist of granular topsoil, skeletal soils, and saline soil, the three most problematic soils. The average temperature for the year is 33 degrees Celsius, but there are only a handful of truly tropical and humid months. It is mild or humid throughout the year with less precipitation. The wettest months are from May to September. In addition, the majority of land uses are agricultural, and the middle and upper portions of the plateau are heavily cultivated with upland commodities such as cassava, sugarcane, corn, rice, and oil palm. Rice fields are expanding in the central portion of the study areas.

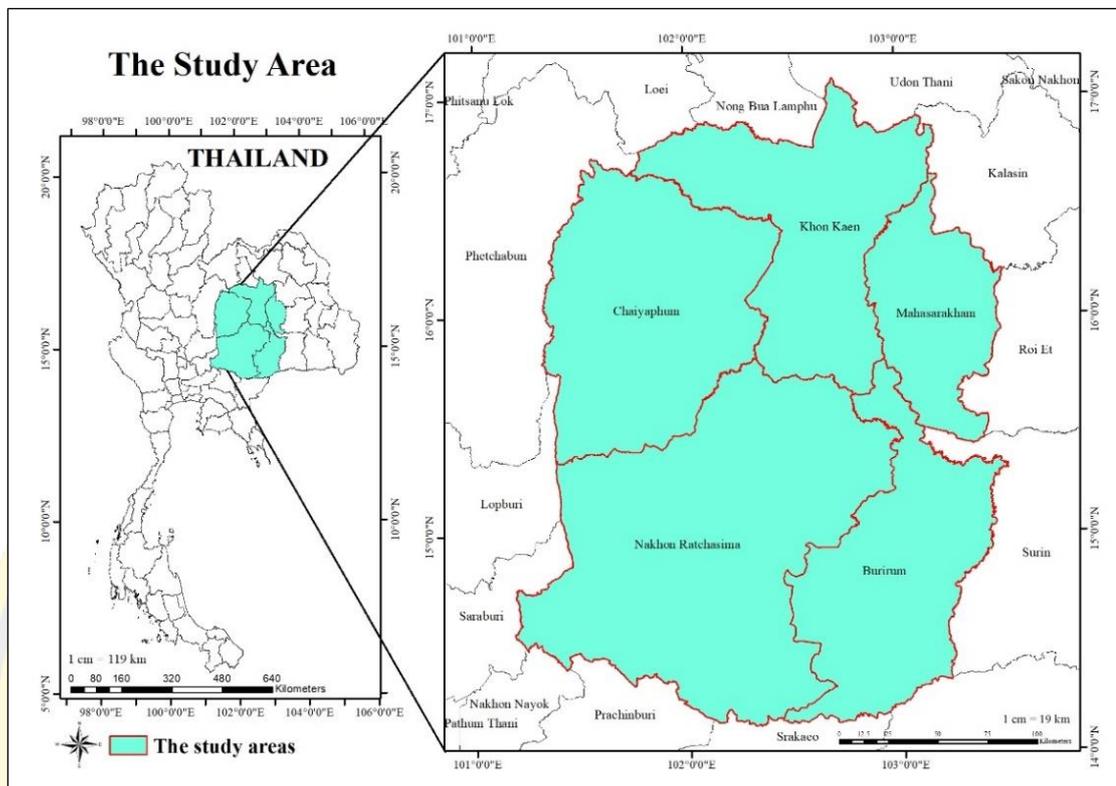


Figure 15 Location of the study area on the overview map of Thailand

Workflow of research

From beginning to completion, the methods for conducting this investigation are described. Figure 17 displays all data variables and sources, including database preprocessing such as soil and vegetation index extraction, backscattering from radar data, topography, the mean value of precipitation, and temperature stacking. While soil sampling data were processed according to soil series and converted from vector to raster formats, the data were also converted from vector to raster formats. Using Pearson correlation analysis, a statistical analysis of soil sampling and the correlation between soil sampling and all data variables were determined. For model prediction, three machine learning algorithms were used to estimate the soil organic carbon storage in the study area to identify the most effective model with the most strongly advantageous feature variables. Finally, soil organic carbon stock in different agricultural land use types was computed using the Zonal analysis for collected soil organic carbon value derived from the best model's output.

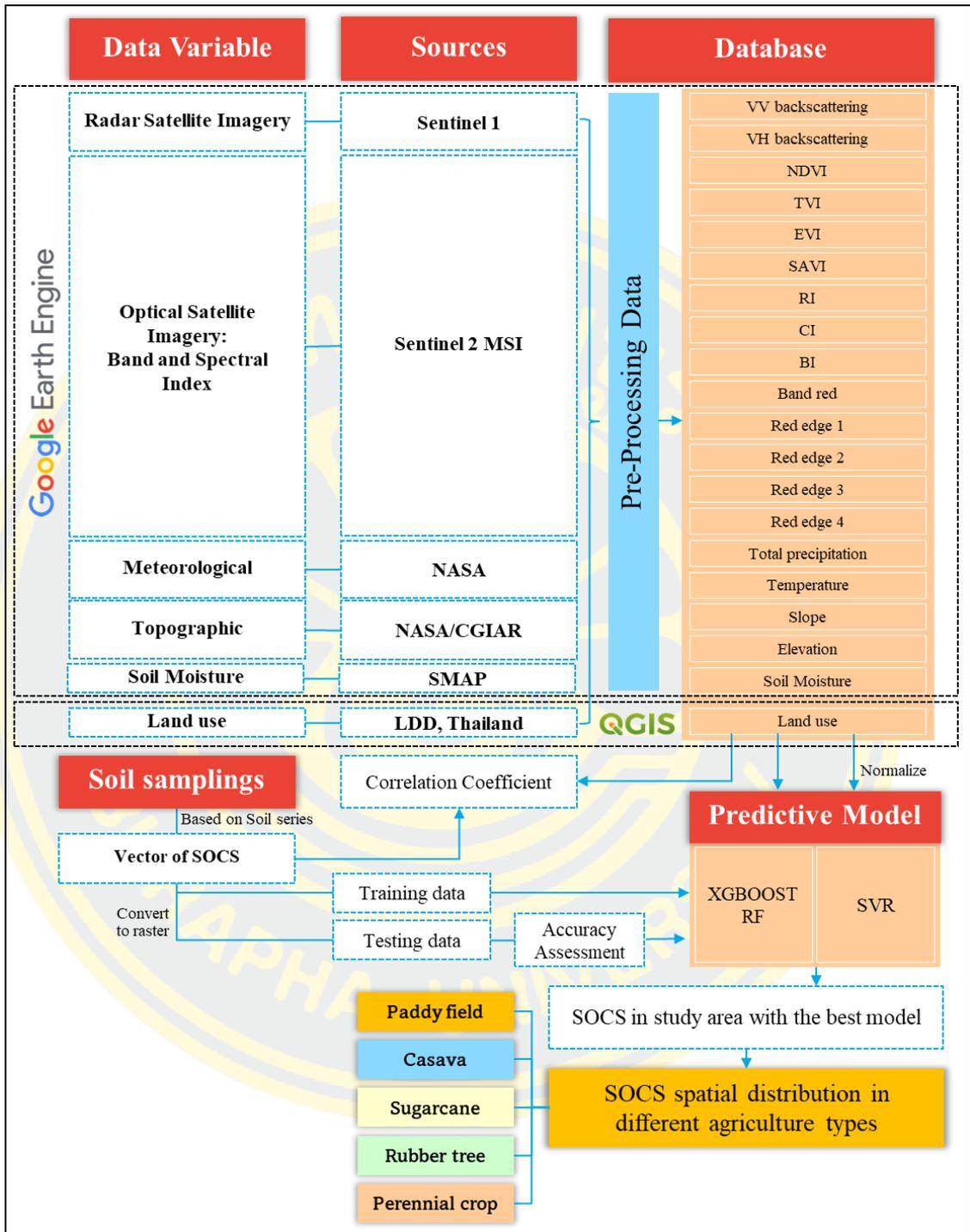


Figure 16 The overall workflow of this study

Soil sampling and laboratory analysis

Soil samples database was obtained from Soil Resources Survey and Research Division, Land Development Department, Thailand. There were 234 samples in 2020 in total by stratified randomization methods according to different land use types in agriculture areas and different soil texture data at 0-30 centimeters. The point information was recorded by a handheld GPS. Moreover, the soil sample was collected in two types. The first soil sample, called soil sample pit, was sealed, and labeled in a soil collection bag and transported to the laboratory for chemical soil properties analysis. The second is undisturbed soil core sampling method.

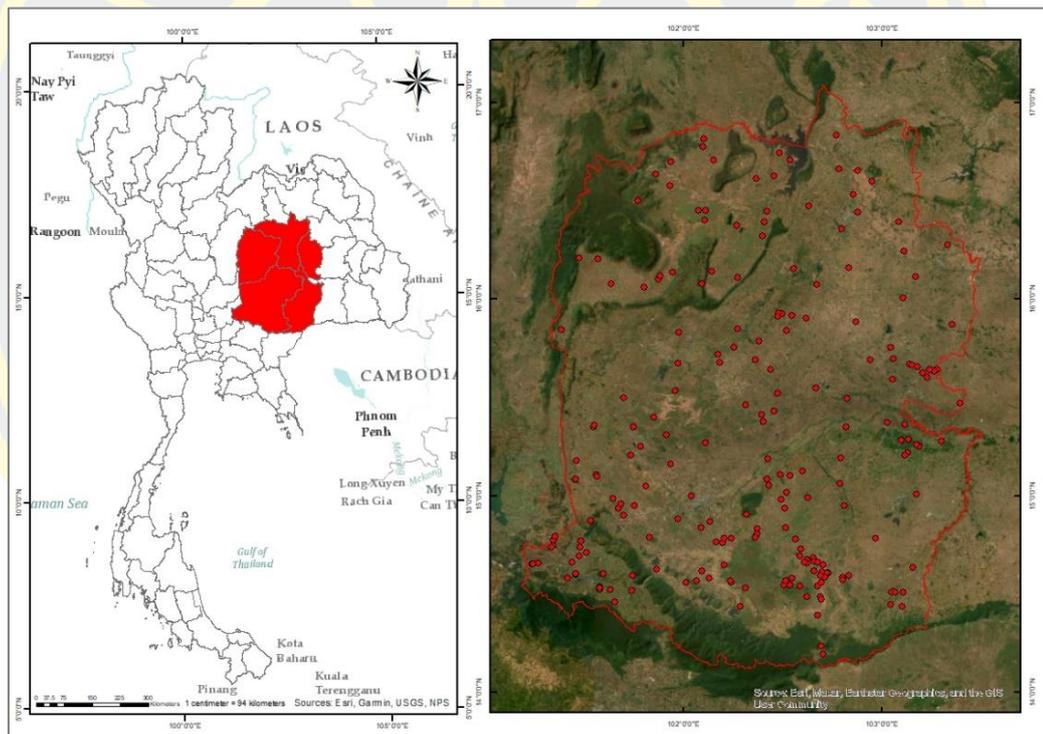


Figure 17 Soil sampling point in the study area (234 soil samplings)

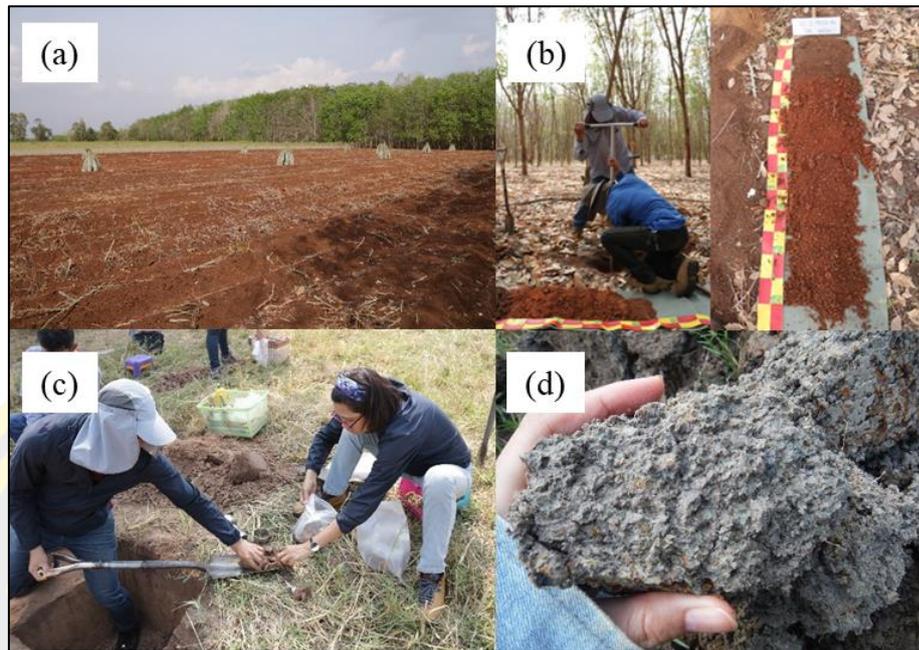


Figure 18 Example of soil pit and soil profile (a) casava field (b) soil profile using auger tools (c) collecting soil sample in the bags (d) characteristics of soil texture in rice fields

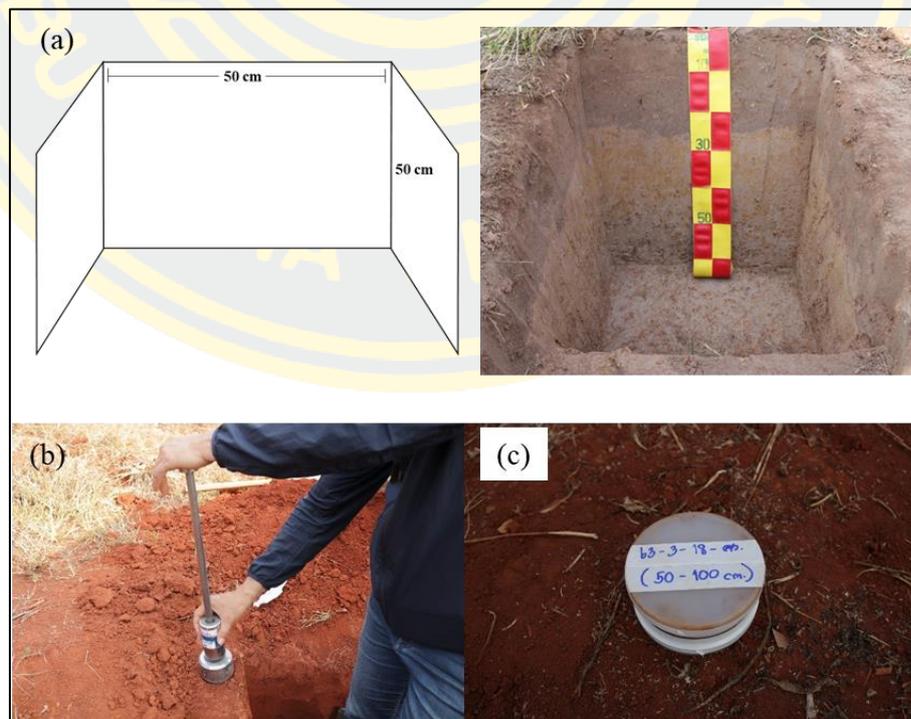


Figure 19 Example of soil collection by undisturbed core method

Physical and chemical analysis

The soil samples that were brought back to the laboratory underwent two steps: preprocessing and chemical testing. First, the soil samples were allowed to dry out in the air; then, the rocks, weed roots, and other debris were removed; finally, the soil was pulverized and sieved through a 2 mm grid to remove any remaining contaminants. In the organic carbon experiments, pretreated soil samples were pretreated by Walkley and Black's Method, mainly acid wet oxidation method. Oxidizable matter in the soil is oxidized by 1 N $K_2Cr_2O_7$ solution. The process that occurs is supported by the heat produced when two volumes of H_2SO_4 and one volume of dichromate are combined. With ferrous sulfate, the residual dichromate is titrated. The value of the titer is inversely proportional to the quantity of carbon in a soil sample. The second soil sample was collected by undisturbed soil core sampling method by a cylindrical core sampler from the soil genesis horizon depths that means the soil retained its natural structure and water content for the total of soil bulk density according to core method (LDD, 2004b).

Particles >2 mm, definitions: Rock and pararock fragments are defined as particles >2 mm in diameter and include all particles with horizontal dimensions less than the size of a pedon. Rock fragments are further defined as strongly cemented or more resistant to rupture; most of these fragments are broken into particles 2 mm or less in diameter during the preparation of samples for particle size analysis in the laboratory. Rock fragments are generally sieved and excluded from most chemical, physical, and mineralogical analyses.



Figure 20 A photograph of a location with numerous gravel and fragments of rock.
 (a) gravel in surface soil (b) remove some huge rock fragments (c) fine and moderately plant root in surface soil, and (d) collect soil and gravel

In this investigation, we utilize particle-size measurement techniques that include laboratory weighing to determine the percentage of gravel in each soil sample. (LDD, 2004a)

$$\text{gravel (\%)} = \frac{\text{Weight of particles } > 2 \text{ mm}}{\text{Total weight of soil sample}} \times 100 \quad (3-1)$$

Finally, soil organic carbon stock (SOCS) was calculated according to GSP Secretariat and ITPS (FAO, 2017) and Land development department in Thailand (Puttaso Annop, 2016), (Duarte et al., 2022) using Equation:

$$\text{SOCS} = D \times BD \times \text{SOC} \times \left(1 - \frac{\text{gravel (\%)}}{100} \right) \quad (3-2)$$

where, SOCS is soil organic carbon stock (ton C/ ha^{-1}), SOC is soil organic carbon concentration (g C kg^{-1}), BD is bulk density (g cm^{-3}), D is depth of horizon (m), gravel (%) is the percentage of gravel in the soil sample, in this study SOCS will use ton C/ha^{-1} unit.

Statistical data analysis of soil organic carbon

Statistical analysis of soil organic carbon stock after data were analyzed from laboratory and calculated soil organic carbon stock. The statistical analysis can be used to preprocess the data before training a machine learning model that can help and understanding data distribution, outlier approach normalizes the data and improves the performance of the machine learning algorithm. The value of statistics was measured the skewness distribution and kurtosis including the maximum, minimum, mean, medium, and Standard Deviation (SD) value using SciPy statistical functions on Visual Code program (Python3). Furthermore, Pearson correlation was used to analyze the correlation between remote sensing features and SOCS to describe the valuable insights into the relationship between these variables.

Acquisition and processing of data

This study used Google Earth Engine and satellite data. Table 4 shows that Google Earth Engine, a cloud-based geospatial data analysis tool, lets users easily access and analyze enormous quantities of satellite imagery, climate data, and other geospatial data. Sentinel-1, VV, and VH polarization modes extracted the backscattering coefficient. Sentinel-2 extracted vegetation and soil spectral data. SMAP, a NASA satellite program, collects high-resolution soil moisture data to study the link between soil moisture and soil organic carbon storage. Finally, the NASA/CGIAR dataset includes environmental variables like temperature and precipitation that affect soil organic carbon storage. The dataset was accessible using Google Earth Engine via the Google collab python API from January 1, 2020, to May 30, 2020, after harvest and drought seasons in the study area. Moreover, Land Development Department in Thailand provided the land use map.

Table 4 List of remote sensing and environment variables

Sources	Data	Feature Collections	Goal	References	
Sentinel 1	Radar data	VV and VH backscattering	Reduces cloud cover and soil moisture	(Huan Wang, 2021)	
Sentinel 2	Red and Red edge bands Vegetation Index Soil Index	B4, B5, B6, B7, B8A	A specific wavelength range can be used to assess soil properties	(Yin Gao et al., 2014)	
		NDVI	Assess plant chlorophyll and vegetation cover.	(Gharahi Ghehi, 2019)	
		TVI	Vegetation cover and process of photosynthesis	(Ali, 2020)	
		EVI	A higher rate of carbon uptake from the atmosphere	(Zhang, 2017)	
		RI	Soil SOC accumulation potential.	(Li, 2016)	
		CI	Degraded into SOC	(Zhao, 2017)	
SMAP	Surface Moisture	SSM	decompose organic materials and generate soil carbon	(Wang, 2019)	
		Topography	Slope Elevation	Impact soil surface erosion and runoff	(Zhang, 2019)
		Meteorological data	Rainfall Temperature	promote plant growth and increase SOC decrease in SOC levels	(Chen, 2019)
LDD	Land use map	Land use	intensive land disturbance and management practices	(EEA, 2022)	

1. Sentinel 1 C-band synthetic aperture radar (SAR)

The data used to calculate the backscattering coefficient for Sentinel-1 came from the European Space Agency and were accessed between January 1 and May 30, 2020. At 5.405GHz (C band), the Sentinel-1A radar operates as a dual-polarization C-band Synthetic Aperture Radar (SAR) device. This collection features the S1 Ground Range Detected (GRD) sceneries obtained using dual-band cross-polarization modes, vertical transmit/horizontal receive, VV and VH in a

two-pass (ascending and descending) configuration, and VV and VH in a vertical transmit/horizontal receive configuration. (Xiaohang Li, 2021). The spatial resolution of the S1 images used was 10 m in the interference width (IW) mode. Utilizing the Google Earth Engine platform and the Google Earth Engine Python API, the procedure for downloading and extraction method was executed. This procedure involves downloading an S1 image from the metadata name "*Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling*" in order to create a homogeneous subset of S1 data that can be filtered using metadata properties. The specific metadata fields used for filtering in this investigation include the following properties:

Table 5 Sentinel-1A data properties for filtering.

Date	Properties	Details
2020-01-01	transmitterReceiverPolarisation:	'VV', 'VH'
to 2020-05-30	instrumentMode:	Interferometric Wide Swath (IW)
	orbitProperties_pass:	'ASCENDING' and 'DESCENDING'
	resolution_meters:	10
	resolution:	'M' (medium)

Finally, all S1 images were calculated the sigma VH: sigma VV ratio in ascending and descending pass, the mean value was computed at the field scale. Then, two S1 image resampled bands for each image at resolution of 30 m, were integrated into an image and clipped by a shapefile of the study area boundary.

2. Sentinel-2 MSI level 2A

The ESA collected the Vegetation and soil spectral Index from Sentinel-2 MSI Level 2A data, which were then converted to surface reflectance data accessible between 1 January 2020 and 30 May 2020. (Castaldi et al., 2019)

Sentinel-2 images contain spectral information between 0.43 and 2.28 microns. Every normal band of the S2 image was downloaded from "*Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*" metadata with a 30 m spatial resolution using Google Earth Engine and the Google Colab Python API with the eemont module. (Montero, 2021) Band were extracted to include in the band calculation and modeling as shown in Table 7. The red and red-edge bands (B4, Red-edge1, Red-edge2, Red-edge3, Red-edge4) derived from Sentinel-2A data were incorporated as model features to improve the accuracy of over the common variable model, as shown in Table 6. (Boqiang Xie, 2022)

Table 6 Sensor spectral and spatial characteristics of the red and red-edge band

Band	Description	Wavelength (nm)	Spatial resolution (m)
B4	Red	664.5/665	10
B5	Red Edge 1	703.9/703.8	20
B6	Red Edge 2	740.2/739.1	20
B7	Red Edge 3	782.5/779.7	20
B8A	Red Edge 4	835.1/833	20

In this investigation, we chose a number of vegetation indices and soil index variables to which SOC is sensitive. The Normalized difference vegetation index (NDVI), the Transformation vegetation index (TVI), the Enhanced vegetation index (EVI), the Redness index (RI), and the color index (CI) were some of the vegetation indices that were chosen. The soil brightness indices (BI) include the Soil Adjusted Vegetation Index (SAVI).

Table 7 Spectral index information for Sentinel-2A data.

Type	Spectral Index	Name	Formula	
Vegetation	NDVI	Normalized Difference	$\frac{NIR - R}{NIR + R}$	
		Vegetation Index	$\frac{NIR - R}{NIR + R}$	
	EVI	Enhanced Vegetation	$2.5 \times \frac{NIR - R}{NIR + 6 \times R - 7.5 \times B + 1}$	
		Index	$2.5 \times \frac{NIR - R}{NIR + 6 \times R - 7.5 \times B + 1}$	
	TVI	Transformed Vegetation Index	$\sqrt{\frac{NIR - R}{NIR + R}} + 0.5 \times 100$	
Soil Index	RI	Redness Index	$\frac{R \times R}{G \times G \times G}$	
		CI	The color index	$\frac{R - G}{R + G}$
			Soil Adjusted	$\frac{(1.0 + L) * (NIR - R)}{(NIR + R + L)}$
	SAVI	Soil Adjusted Vegetation Index	$\frac{(1.0 + L) * (NIR - R)}{(NIR + R + L)}$	
	BI	Bare Soil Index	$\frac{\sqrt{(R \times R) + (G \times G)}}{2}$	

3. Topography

The topographic data were obtained from the digital elevation dataset provided by the Shuttle Radar Topography Mission (SRTM), which was initially developed with the intention of providing consistent and high-quality elevation data on a scale that was nearly global in scope. This particular iteration of the SRTM digital elevation data has been processed to remove any gaps in the data as well as make its use more straightforward. The SRTM Digital Elevation Data Version 4 dataset was downloaded from the NASA/CGIAR in Google Earth Engine platform. This dataset, *ee.Image("CGIAR/SRTM90_V4")*, covered the research area with a resolution of 90 meters, with elevation band and it was utilized to extract a variety of topographical characteristics. The elevation variables that were used in this study were calculated slope using Google Earth Engine python API, cropped to the study area boundary and finally resampled to 30 m using the NEAREST method by QGIS software. (Tao Zhou, 2020).

4. Meteorological data

The meteorological data that are provided by GLDAS-2.1 have a historically consistent series from 1948 to 2014, and they combine model and observational data from 2000 to the present at a resolution of 27.83 kilometers. The total precipitation rate and air temperature data for the climate parameters were obtained from the 0-0.1 kg/m²/s and 206.8-327.66 K (estimated min or max value) datasets from NASA GES DISC at NASA Goddard Space Flight Center by *ee.ImageCollection("NASA/GLDAS/V021/NOAH/G025/T3H")*. These datasets were accessed on 1 January – 30 May 2020. We calculated the total precipitation rate using the *"Rainf_f_tavg"* function, and the air temperature using the *"Tair_f_inst"* function. The total precipitation rate and air temperature data were resampled to resolutions of 30 meters using the resampling function in the QGIS program. This provided a high-quality and straightforward database that could be used to evaluate the precision of the climate variables used in this investigation.

5. Surface soil moisture

The updated two-layer Palmer model creates soil moisture from satellite-derived SMAP Level 3 soil moisture readings using a 1-D Ensemble Kalman Filter (EnKF) data assimilation technique. The day's climatology determined soil moisture anomalies. Estimating climatology using the complete SMAP satellite observation and 31-day-centered moving-window approach. SMAP observations improve model-based soil moisture forecasts in poorly instrumented regions with weak precipitation data. Retrieved soil moisture estimate from the disaggregated/downscaled vertical polarization brightness temperature at 9-km grid cell; AM overpass with resolution at 9000 meters. We were accessed on 1 January – 30 May 2020 using *"ee.ImageCollection("NASA/SMAP/SPL3SMP_E/005")"*, provided from Google and NSIDC. We calculated the mean soil moisture with units at 1 (volume fraction) using the *"soil_moisture_am"* band. Finally, data were resampled to resolutions of 30 meters using the resampling function in the QGIS program.

6. Land use mapping

Because of its ability to store water and nutrients and offer other ecosystem services, land use change has the potential to have an impact on the functions of the soil that are either directly or indirectly related to the organic matter and carbon stock of the soil. In this investigation, we take advantage of the Land use map from the Land Development Department in 2020, complete with its properties in agricultural regions and presented in shapefile format at a scale of 1:25,000. Land use level 3 was selected in agriculture land including paddy field, crop, perennial, orchard, horticulture, and etc. Using QGIS software converted land use vector to raster using “classes” attribute and resampling image to 30 meters resolution.

Table 8 Land use classification in the study area

LU_DES	LUL1	LUL2	LUL3	classes	Area (ha ⁻¹)	%
Water	W	-	-	1	208,317.21	3.48
Urban	U	-	-	2	414,319.43	6.93
Miscellaneous	M	-	-	3	276,217.85	4.62
Forest	F	-	-	4	860,669.03	14.40
Paddy field	A	A1	A101	5	2,065,772.48	34.56
Corn	A	A2	A202	6	87,689.66	1.47
Sugarcane	A	A2	A203	7	750,924.29	12.56
Casava	A	A2	A204	8	804,907.12	13.46
Para rubber	A	A3	A302	9	129,120.46	2.16
Oil palm	A	A3	A303	10	5,477.56	0.09
Perennial crop	A	A3	-	11	114,900.75	1.92
Orchart	A	A4	-	12	47,617.77	0.80
Mixed field crop	A	A2	-	13	21,885.61	0.37
Others crop	A	-	-	14	190,180.79	3.18
Total					5,978,000.00	100

Predictive model

The machine learning model is a relatively mature model. In this study, five ML algorithms, including random forest (RF), extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR) were chosen to predict and estimate soil organic carbon stock, which can obtain effective results. This study uses the GPU-accelerated cuML library, which implements machine learning algorithms quickly and efficiently. We also calculated feature importance from each model using each machine learning method's permutation importance calculation to rank each variable's relevance. The permutation importance method was utilized within the cuML environment, permutation, which utilized Python 3.10, to compute the important values for each feature through the cuML's single GPU estimators. All of model, we use Visual Studio Code python 3.10. Table 9 shows a brief description of the machine learning used in this study presented as follows:

Table 9 Three predictive models selected in this study.

Machine Learning Model	References
RF	(Xiaohang Li, 2021)
XGBoost	(Boqiang Xie, 2022)
SVR	(Li, 2021)

1. Random forest regression (RF) algorithms

The cuML Random Forest algorithm runs much faster than its CPU-based counterpart, especially on large datasets as this study. This is because the algorithm takes advantage of the parallel processing capabilities of the GPU to speed up computations. The RF algorithm had the lowest average MAE value (0.17), as well as the lowest RMSE value (0.20). (John et al., 2020) In each estimator of RF is trained on a different bootstrap sample having the same size as the training set that can be found in the *cuml.ensemble.RandomForestRegressor*. In cuML-RF, users typically modify two hyperparameters to control the complexity of the models: (a) the number of trees (or iterations) (*ntree*), which also corresponds to the number of decision trees;

random forests will overfit if the number is too large; and (b) mtry, which represents the number of indicators that are randomly sampled as candidates at each split. In this case study, we will adjust the ntree and mtry parameters, which have the following influence on our random forest model.

In this study, the model performance is obtained from each combination of the hyperparameters tuning with the grid search method. During the process of tuning RF hyperparameters, this was done to eliminate any potential for bias in the data selection. Moreover, Tree nodes between 400 and 1000 was probed, with 400 tree nodes the R^2 value was low and the RMSE is not stable. With 1000 tree nodes a highest value R^2 was obtained with a stable RMSE, with 500 tree nodes the results were the same, so in order to optimize the model on this study 500 tree nodes were used. (Wang et al., 2021)

2. Extreme gradient boosting (XGBoost) algorithms

Extreme gradient boosting (XGBoost) is a gradient boosting algorithm extensively utilized in regression tasks. XGBoost increases the efficiency of model learning by utilizing parallel computation and an additive decision tree training strategy to transform multiple weak learners into strong learners. Using this method, XGBoost can perform regression assignments. Unlike previous decision tree algorithms, XGBoost is based on a second-order Taylor formula expansion, incorporates a regularization module, and predicts through a number of additive functions, thereby effectively controlling the overfitting phenomenon. XGBoost algorithm was implemented using the Sklearn machine learning library. Using 10-fold cross-validation and grid search, crucial hyperparameters such as the number of trees, maximum tree depth, maximum number of nodes or leaves in the decision tree, and learning rate are optimized to determine the optimal model parameters, as seen in Xie et al. (2022), Wang et al. (2021) In the XGBoost algorithm, the contribution of the variable through each tree model is determined by calculating the gain as follows:

$$\text{gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (3-3)$$

where the tree model's left leaf is represented by G_L and H_L . $\frac{G_L^2}{H_L+\lambda}$ indicate left subtree scores. G_R and H_R are related to the correct branch of the tree model.. $\frac{G_R^2}{H_R+\lambda}$ indicates correct subtree ratings. $\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}$ represents the unbroken node score. Regularization parameters govern tree structure simplicity to prevent overfitting. Variables with greater gain values are more relevant for model prediction.

In this study, the model performance is derived from each combination of the hyperparameters tuning with cross-validation (CV) methods in order to improve the accuracy of the samples and to extract the respective characteristics of each sample. K-fold CV is one of the most commonly used CV methods in machine learning, but there is no rule for determining the value of k.

3. Support vector regression (SVR) algorithms

SVR regression. Classification SVMs handle categorical outputs, but this variant handles continuous numerical outputs. SVR works well for complex, high-dimensional datasets. SVR is a regression algorithm. It finds a hyperplane in a high-dimensional feature space that optimizes the margin between expected and actual training data outputs. The model is trained to reduce the error between expected and actual outputs, subject to a margin restriction, utilizing support vectors to form this hyperplane. SVR performance depends on the kernel function, which transforms input into a high-dimensional feature space for hyperplane separation. (Chih-Chung Chang, 2022) Unlike other regression algorithms, SVR can handle non-linear relationships and datasets with multiple characteristics without overfitting. SVR model stored soil organic carbon least. SVR model soil organic carbon storage was lowest. The support vector regression model also had the lowest root mean squared error (RMSE) and the highest R^2 values (14.9 Mg C ha⁻¹ and 0.6, respectively); therefore, it was the most effective method for predicting SOC inventories.(Were et al., 2015) In this study, we utilized SVR in TensorFlow that provides a number of built-in kernel functions. a linear regression layer on top of the kernelized input data to perform the regression.

Accuracy assessment

In this study, the estimating of SOC stock is studied, and RF, SVR, and XGBoost were constructed by machine learning methods. The modeling set of model evaluation, we selected 70 % soil organic carbon data for training, and verified the remaining 30% data. Based on this, we selected three prediction indexes to determine the prediction effect of each index, mainly including R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) and to determine the prediction accuracy of the data.(Tao Zhou, 2020) When R^2 gets closer to 1, the model has an excellent fitting degree and remarkable stability. In a similar vein, the model is more predictive and robust than before, with a smaller MAE and RMSE. To summarize, the formula is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (3-4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (3-5)$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (P_i - \bar{O}_i)^2} \quad (3-6)$$

Where n represents the number of samples; P_i and O_i represent the predicted and observed values at site i , respectively.

Soil organic carbon spatial distribution and validation

After effectively training the soil organic carbon storage models with the best model and significant factors, their accuracy was evaluated using R^2 , Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) for the study area. The model was then used to calculate soil organic carbon storage values for each pixel using crop spatial raster data with a shapefile in Python according to agricultural land use types by mean, minimum, and maximum values of SOCS in the study area and to visualize the prediction. The final method generates a soil organic carbon storage map for the study area and computes carbon stock in different land use types in agriculture, including rice, sugarcane, cassava, rubber tree, and perennial crops: a Shapefile of Land use types in agriculture in 2020 from the Land Development Department in

Thailand, at a scale of 1:25,000, containing all land use types. In this investigation, we chose five types of agricultural regions that produce rice, sugarcane, cassava, rubber tree, and perennial trees. Table 10 shows eight levels of soil organic carbon level based on the mean value. Classifying soil organic carbon concentration into levels helps explain its distribution and fluctuation. QGIS 10.6 displays this method.

Table 10 The levels of SOC stock in this study.

Level	Soil organic carbon stock (ton C/ha ⁻¹)
1	0-5
2	5-10
3	10-15
4	15-20
5	20-25
6	25-30
7	30-35
8	>35

To validate the accuracy of predicting soil organic carbon storage (SOCS) in different crop types and soil sampling locations, the R² (R-squared), RMSE (Root Mean Square Error), and MAE (Mean Absolute Error) metrics are employed. These metrics are used to evaluate the accuracy of the best machine learning model in predicting SOCS, both in terms of individual crop types and soil sampling values.

CHAPTER 4

RESULT

Statistical descriptions

1. Statistics of soil sampling

The statistical values of the training and validation datasets for SOC prediction are shown in Table 9 based on the 234 soil samples obtained in the study area. These values were determined using the training data for the model. The maximal value of the complete SOC dataset was 60.47, the mean was 24.13, the median was 21.34, and the standard deviation (SD) was 14.65. The range of the dataset was 57.55. (Figure 21) The bar chart presents an illustration of the histogram of frequency for the dataset, which has a skewness of 0.56, which shows that the distribution of the dataset is somewhat right skewed. What this indicates is that a greater proportion of its data points have low values for the soil organic carbon storage variable. Therefore, if we train our model on these data, it will have a superior performance when predicting the SOCS of carbon with a lower value as opposed to those with a larger value. In addition, as compared to a normal distribution, a kurtosis score of -0.65 implies a distribution that is somewhat flatter.

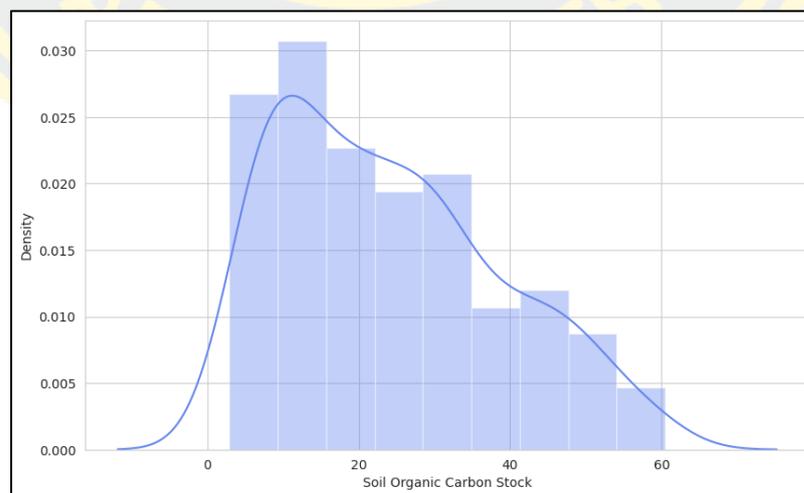


Figure 21 The skewness histogram of soil organic carbon data

Table 11 Descriptive statistics for the entire SOC stock dataset.

SOCS	N	Mean	Min	Max	Median	SD	Range	Skewness	Kurtosis
Total	234	24.13	2.92	60.47	21.34	14.65	57.55	0.56	-0.65
Paddy field	103	17.69	2.94	52.98	15.31	12.91	50.04	1.09	0.73
Cassava	41	25.03	7.27	57.89	22.28	13.45	50.62	0.68	-0.52
Sugarcane	52	22.84	4.26	60.47	19.44	15.00	56.21	0.50	-0.82
Rubber trees	16	28.92	6.57	59.87	27.46	15.37	53.30	0.47	-0.41
Perennial crops	22	32.31	4.91	58.9	37.60	17.83	53.99	-0.24	-1.39

N = number; Min = minimum; Max = maximum; SD = standard deviation

2. Statistics of remote sensing data variables

The statistical analysis of remote sensing and environmental variables reveals diverse characteristics across the dataset. For remote sensing imagery, the VV and VH backscatter coefficients have negative means (-2.84 dB and -3.39 dB, respectively) with moderate variability, suggesting symmetrical distributions. Reflectance values for spectral bands (Red to RE4) range from 0 to near 1, with increasing means from 0.10 to 0.21 and slight negative skewness, indicating minimal deviations from normality. Vegetation indices such as NDVI, SAVI, and EVI exhibit positive means (0.24, 0.13, and 0.15, respectively), with NDVI and EVI showing slight positive skewness. TVI has the highest mean (0.90) and kurtosis (1.90), reflecting a narrower range of high values. Other indices, such as BI and CI, display lower means (0.17 and 0.11) and balanced distributions.

Terrain covariates show contrasting variability. Slope is mostly flat (mean = 1.80°) with occasional steep areas (max = 66.72°) and high skewness (4.52). Elevation varies significantly (mean = 189.67 m) with positive skewness (1.79), indicating some higher elevations. Meteorological variables have narrow ranges, with stable temperature (mean = 27.66°C) and low rainfall (mean = $0.001 \text{ kg/m}^2/\text{s}$) showing slight positive skewness.

Surface soil moisture (SSM) ranges from 0 to 7.756, with a mean of 2.66 and moderate variability, indicating localized higher moisture levels. Land use spans 0 to 13 categories, with a mean of 3.89 and moderate variability (SD = 3.03), showing

a slight skew toward lower classifications. This diversity makes the dataset well-suited for modeling relationships between environmental factors and soil organic carbon dynamics.

Table 12 Descriptive statistical analysis of remote sensing and environmental variables

Dataset	Unit	Min	Max	Mean	SD	Skewness	Kurtosis
Remote Sensing imagery							
VV	dB	-5.30	0.34	-2.84	1.53	-0.89	-0.98
VH	dB	-4.99	-1.79	-3.39	1.01	-0.84	-1.02
Red	DN	0.00	0.94	0.10	0.08	-0.14	-1.24
RE1	DN	0.00	0.96	0.13	0.09	-0.39	-1.19
RE2	DN	0.00	0.93	0.17	0.11	-1.72	-1.08
RE3	DN	0.00	0.95	0.19	0.12	-0.74	-1.08
RE4	DN	0.00	0.97	0.21	0.14	-1.75	-1.10
NDVI	DN	-0.55	0.89	0.24	0.20	0.66	0.14
SAVI	DN	-0.25	0.77	0.13	0.10	0.39	-0.15
TVI	DN	0.19	1.18	0.90	0.09	0.54	1.90
EVI	DN	-0.24	0.94	0.15	0.12	0.59	0.22
RI	DN	0.00	171.35	9.10	6.19	-0.22	0.02
BI	DN	0.00	0.38	0.17	0.06	-0.69	-0.22
CI	DN	0.00	0.58	0.11	0.05	-0.18	0.22
Terrain-based covariates							
Slope	Degree	0.00	66.72	1.80	3.775	4.52	25.29
Elev	M	0.00	1348	189.67	178.71	1.79	4.95
Meteorological-based covariates							
Temp	Degree Celsius	26.5	28.20	27.66	0.23	-0.98	-0.97
Rainfall	kg/m ² /s	0.00	0.003	0.001	0.001	0.19	0.80
Surface Soil Moisture							
SSM	1 (volume fraction)	0.00	7.756	2.66	1.78	-0.41	-0.89
Land use types							
Land use	DN	0	13	3.89	3.03	0.193	-0.72

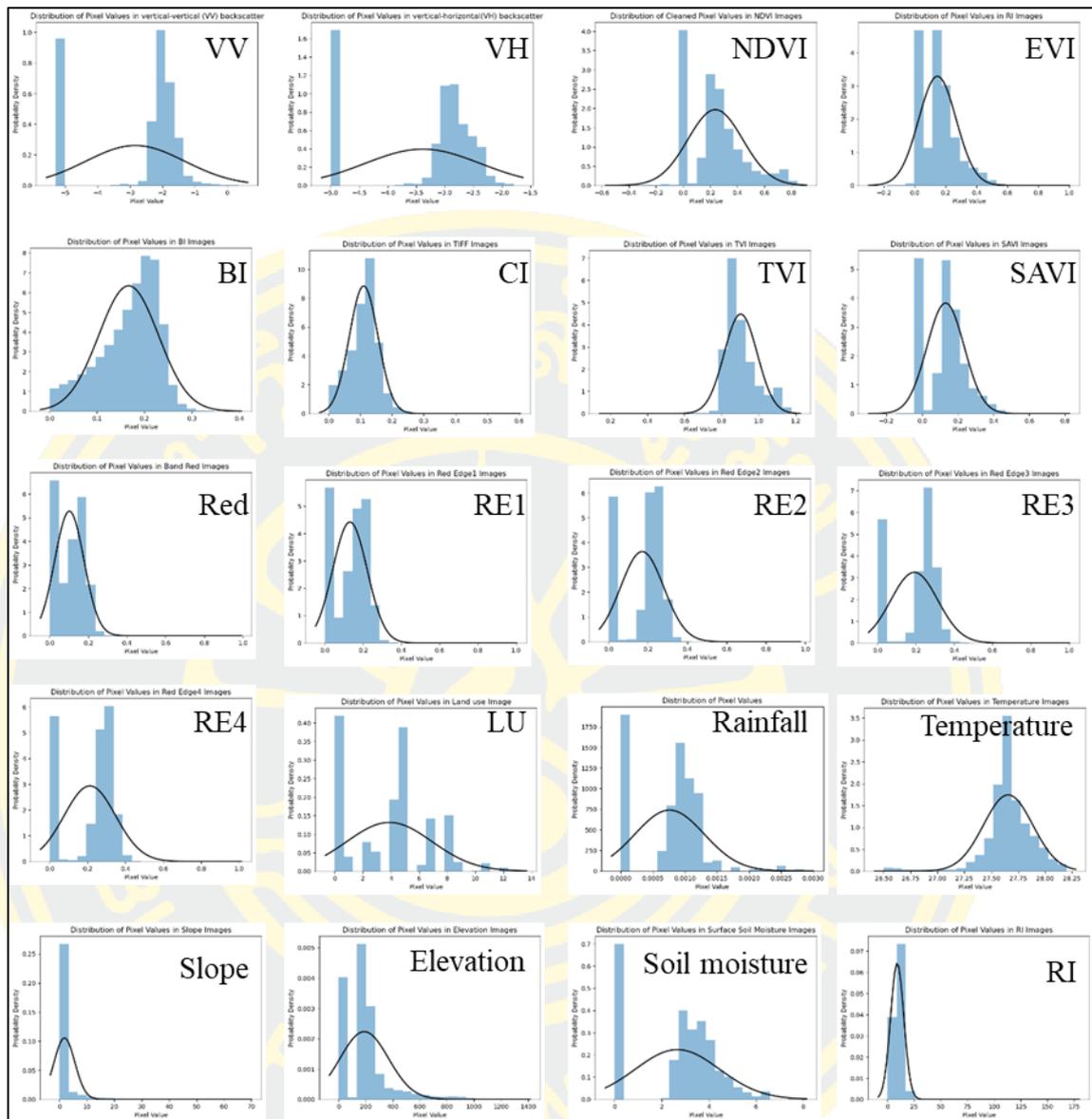


Figure 22 The histogram of remote sensing and environmental variables data

Pearson analysis of the coefficient of correlation between the observed SOCS and the data gathered through remote sensing as well as the environmental factors sampled at various sites. The results show that NDVI, TVI, VH, SAVI, EVI, LU, and Elev had significant with SOCS in the study area with 0.61, 0.6, 0.59, 0.57, 0.58, 0.39 and 0.27, respectively. Consequently, the stock of SOC had a negative correlation with BI ($r = -0.57$), CI ($r = -0.51$), Red ($r = -0.51$), RE1 ($r = -0.42$), and RI ($r = -0.14$). Surprisingly, the SOCS had a significant correlation with

the factors selected that were connected to remote sensing, but the land use variables only had a modest significance with the climate variable. Therefore, characteristics derived from remote sensing were evaluated for their ability to serve as accurate predictors in this study.

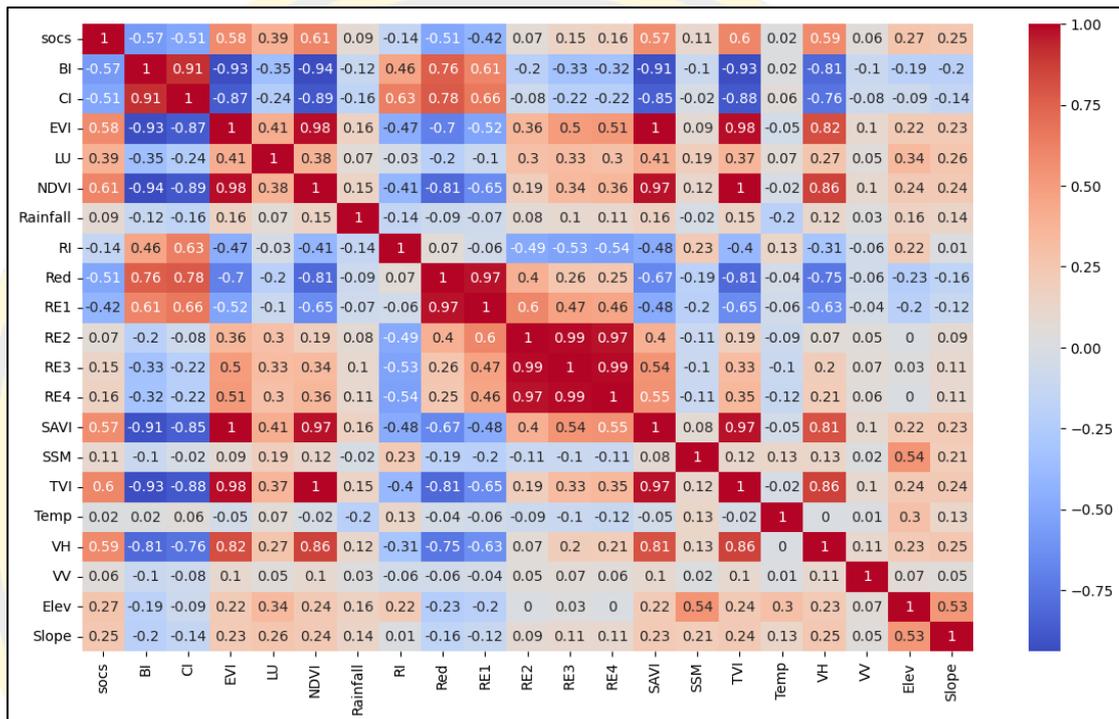


Figure 23 Heatmap of feature collection and soil organic carbon storage using Pearson coefficient analysis.

The importance of remote sensing data and environment variables

Twenty different features of remote sensing information were used as feature variables to arrive at an accurate estimate of the amount of soil organic carbon that was present at the research locations. We examined the degree to which the modeled variables influenced SOCS forecasts in the study region and then standardized the importance of the components to a value of 100% so that comparisons could be made more easily. This was done using a permutation importance calculation. As a result, the comparability of the variables was significantly enhanced, and correlations were discovered between all the selected link

variables and SOCS. Figure 24 illustrates the prioritization and use of three models that were utilized to investigate the influence of these variables on SOCS.

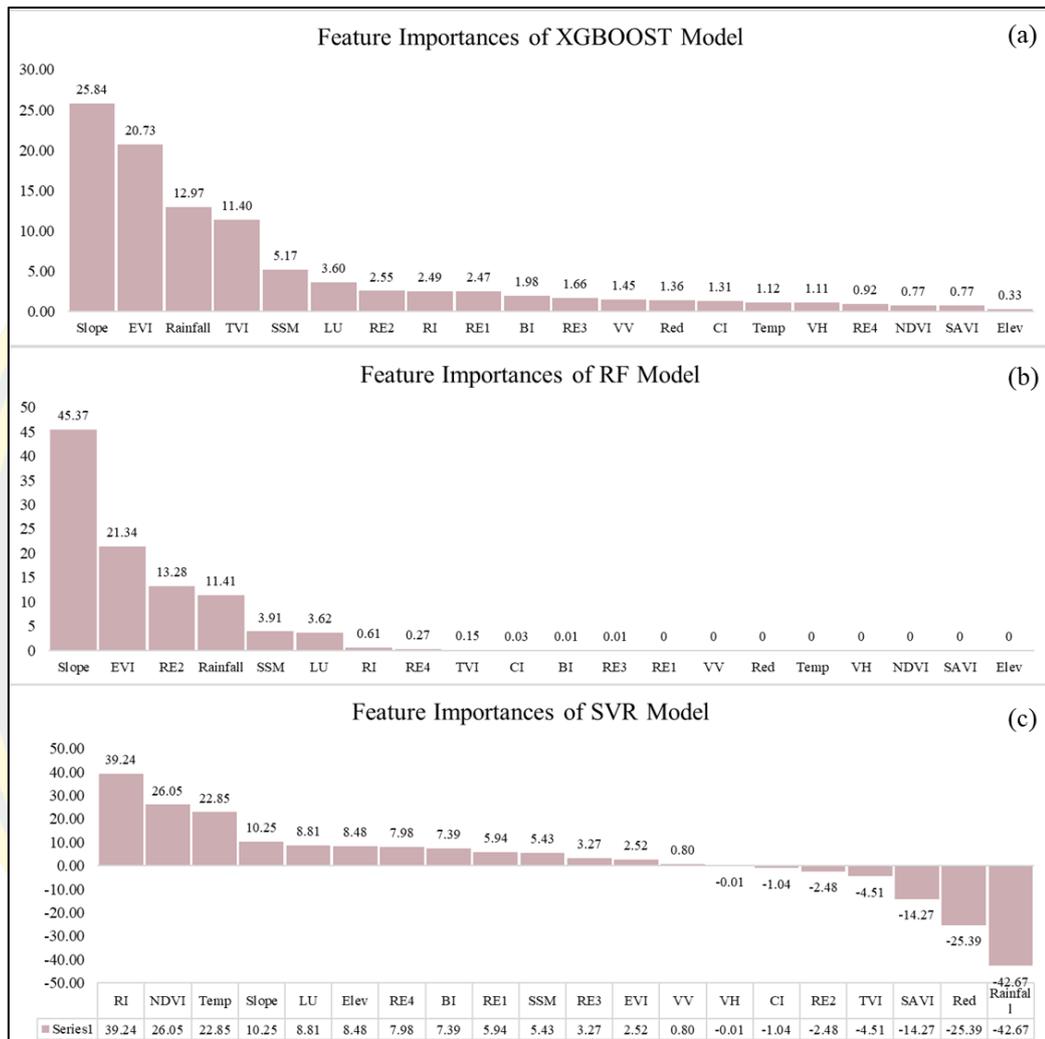


Figure 24 Permutation importance of each models (a) XGBoost model; (b) RF model; and (c) SVR model

According to the bar chart that can be found in Figure 24a, the most important aspect of the XGBoost model is the slope, which is responsible for 25.84% of the total feature importance. The EVI is the second most important element, accounting for 20.73 % of the model's total contribution. The remaining factors, which include precipitation, TVI, soil moisture, and land use, each have their own unique relevance value, which ranges from 3.60 % to 12.97 %. In a similar manner,

the Random Forest model demonstrated that slope and EVI were the characteristics that had the highest feature relevance, with 45.37% and 21.34%, respectively. The second significant number is called RE2, and it takes into consideration precipitation, the moisture content of the soil, and land usage in the proportions of 13.28, 11.41, and 3.91% respectively.

In contrast to the SVR model, it was found that the RI variable (39.24%) was the most relevant part of the remote sensing data. Following NDVI, temperature (22.85%) and slope (10.25%) are the characteristics that are the next most important after NDVI (26.05%). Surprisingly, the SVR model came out with negative feature importance values for seven features that is rainfall, red band, SAVI, TVI, RE2, CI, and VH with at feature importance values at -42.67, -25.39, -14.27, -4.51, -2.48, -1.04, and -0.01, respectively. The negative feature importance value illustrates that the values of those features increase, the predicted SOCS value tends to decrease. Moreover, negative feature importance values can provide insights into the relationship between the features and the target variable in SVR model. Although it might seem counterintuitive at first, it simply means that those features have a negative correlation or influence on the SOCS in the study area.

In conclusion, the XGBoost and Random Forest models arrived at the conclusion that slope and EVI are the characteristics that are the most significant. The other factors, such as rainfall, TVI, soil moisture, and land use, each played a different but significant role in the model. Seven features of the SVR model had significant negative values. The RI variable was shown to be the most relevant piece of information for the remote sensing data in the study, followed by the NDVI, temperature, and slope.

Performances of machine learning models

In this study, soil organic carbon was predicted using 30 m resolution on the basis of Remote sensing data, topography, climate, and land use data. Four models of machine learning are RF, XGBOOST, SVR, and RF. Each model's appropriate parameters are enumerated in Table 13. When using 10-fold cross-validation, XGBOOST provides the most accurate predictions. When `n_estimate` is set to 100 and `max_feature` is set to 10, RF generates the good accurate forecast. In addition,

following the normalization of the data and the running of the model with the `n_estimation` parameter set to 100, we found that SVR provided the most accurate forecast.

Table 13 The hyperparameters of the three models with the highest performance

Model	Model Parameters
XGBOOST	Num_fold =10, Num_boost_round = 100, tree_method = gpu_hist
RF	tree nodes = 500, Max_features = 10, n_bins = 16
SVR	epochs = 10

The predictive abilities of three machine learning algorithms—RF, SVR, and XGBOOST—were analyzed and compared to investigate the impact that a wide variety of variable combinations had on the outcomes of SOC prediction. Because each variable contributes something beneficial to the prediction model for SOCS, utilizing all the model's variables will result in a greater degree of accuracy in the model's output. When modeling and forecasting, we considered all the variables. The ability of the three models to make accurate forecasts is compared in Table 6, which may be seen here. The R^2 , RMSE, and MAE values, in addition to the relative relevance of each feature, varied greatly between the various models.

Table 14 Comparison and evaluation of predicted model performance.

Model	R^2	RMSE	MAE
XGBOOST	0.803	5.569	3.204
RF	0.64	7.53	4.57
SVR	0.59	8.11	5.11

The evaluation of the three models' respective capacities for making accurate forecasts is presented in Table 6. At a resolution of 30 meters, different models

produced varying results when combined with data from remote sensing and environmental variables; nevertheless, the overall precision of the XGBOOST model is superior with $R^2 = 0.803$, RMSE = 5.569, and MAE = 3.204. In comparison, the RF model delivered a yield with $R^2 = 0.58$, RMSE = 8.10, and MAE = 5.11, while the SVR model performed well with $R^2 = 0.48$, RMSE = 8.06, and MAE = 5.17. Even though we utilized the same feature variables, the prediction effect was observed. Using the XGBOOST model with twenty feature variables, Figure 28 displays the SOCS in the study area.

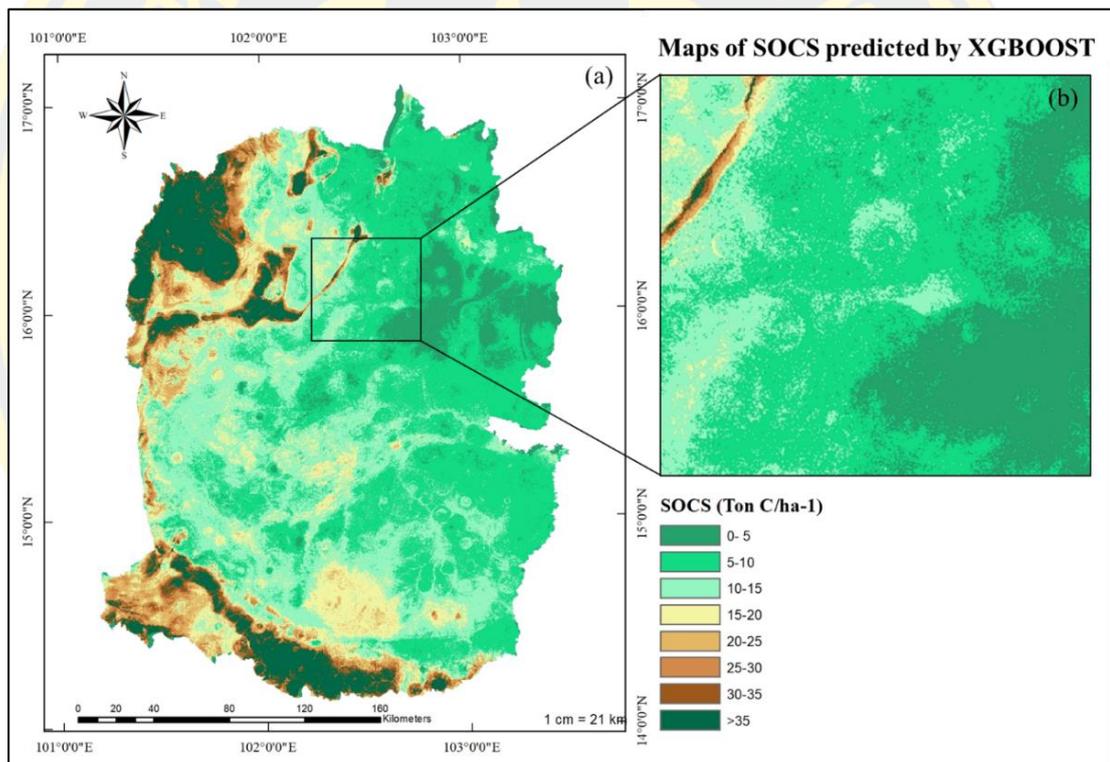


Figure 25 SOCS at 0–30 cm predicted from the XGBoost model

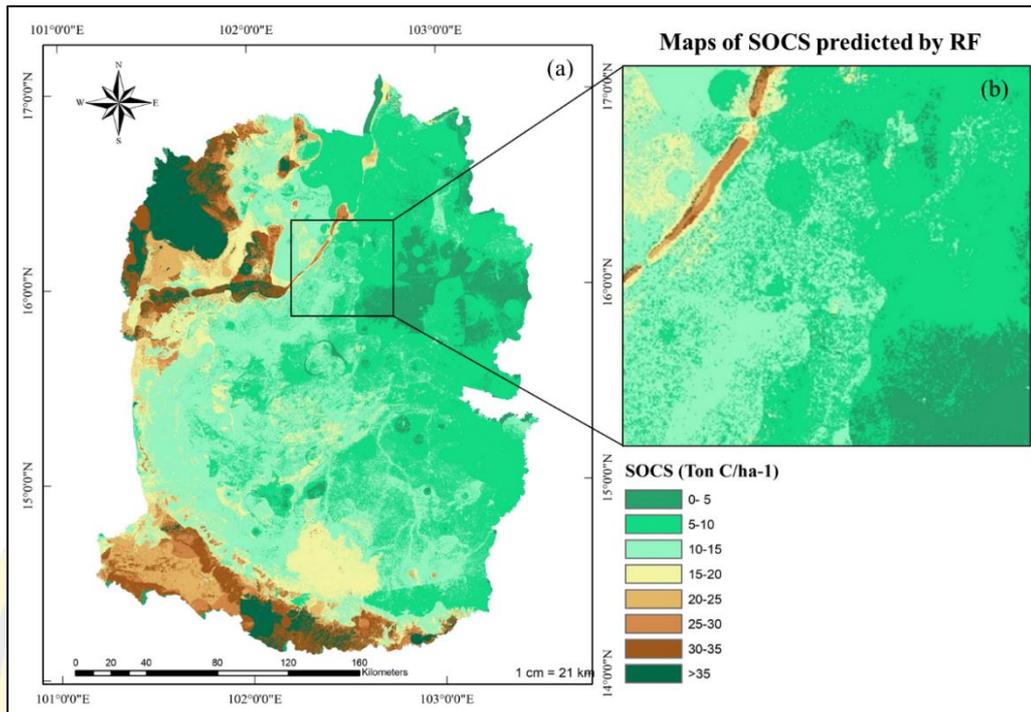


Figure 26 SOCS at 0–30 cm predicted from the Random Forest model

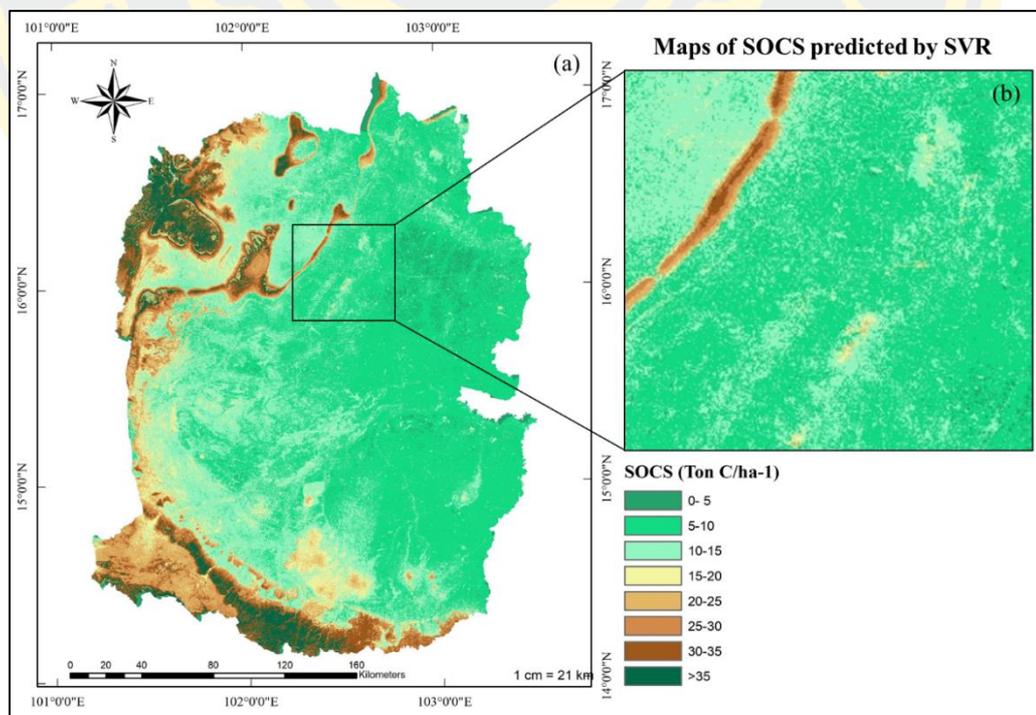


Figure 27 SOCS at 0–30 cm predicted from SVR model

Soil organic carbon spatial distributions and validation

The lower part of northeastern Thailand is comprised of a variety of minerals, landscapes, land use, and soil types, as well as flat plains, undulating hills, and rugged mountains. Soil originates primarily from sandstone and shale. As a result, the majority of the soils in the study area were infertile soils with low clay mineral content and low fertility, resulting in negative carbon accumulation. The average temperature ranges from 32.6°C to 22.2°C, and there is 1,412.4 mm of precipitation. Most of the agricultural land is paddy fields, cassava, sugarcane, rubber tree, and perennial crop, with 2,065,772.48 hectares, 804,907.12 hectares, 750,924.29 hectares, and 129,120.46 hectares, respectively.

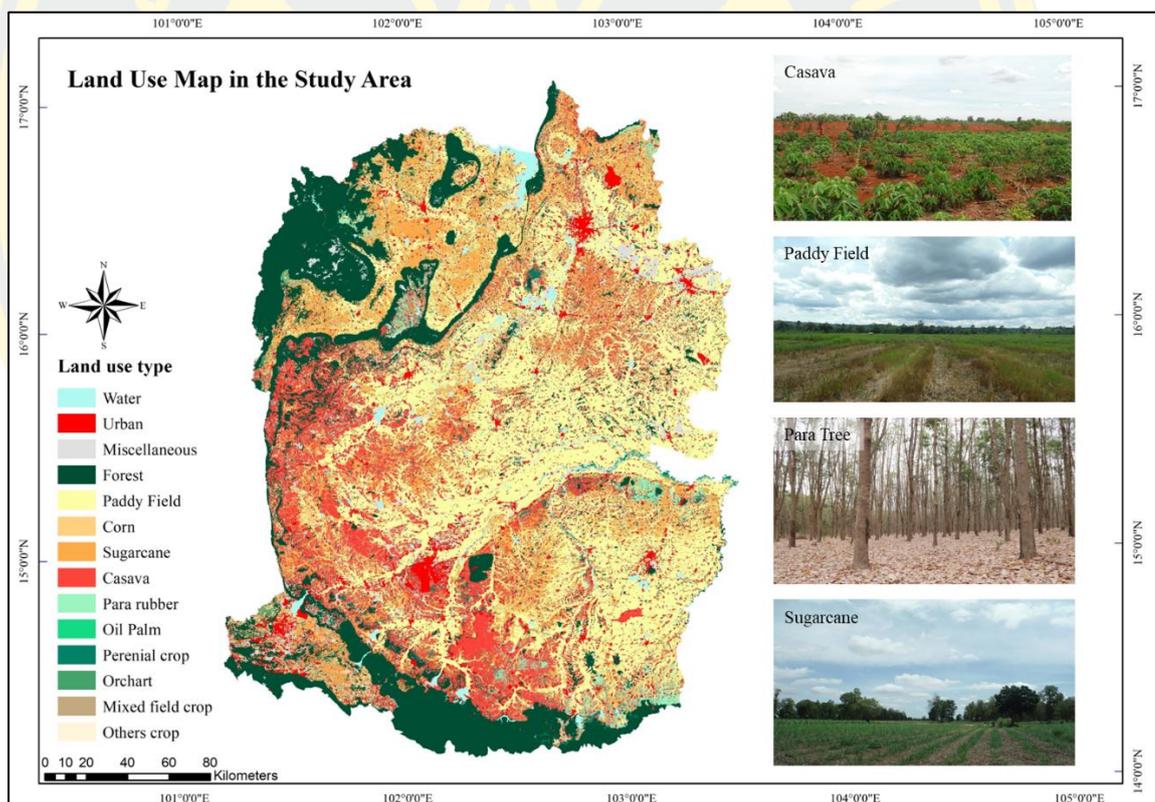


Figure 28 Land use type in agricultural area in the study area

The XGBOOST model has the highest prediction accuracy, show as Table 6. We selected XGBOOST model to collect the value of soil organic carbon stock in rice, cassava, sugarcane, rubber tree and perennial crop, show as Figure 31, using crop

spatial raster data with a shapefile in Python according to agricultural land use types. The results indicate that the range of soil organic carbon stock in paddy fields is between 0.03 to 42.49 tons C/ha⁻¹, with a mean value of 9.86 tons C/ha⁻¹. SOCS at Casava ranges from 0.04 to 50.16 tons C/ha⁻¹, with a mean value of 12.58 tons C/ha⁻¹. SOCS ranges from 0.01 to 47.78 tons C/ha⁻¹ for sugarcane. Rubber tree SOCS ranges from 0.09 to 46.67 tons C/ha⁻¹, with a mean of 14.29 tons C/ha⁻¹. Moreover, the SOCS range for perennial crops is 0.02 to 44.41 tons C/ha⁻¹, with a mean value of 10.72 tons C/ha⁻¹, shown as Table 14.

Table 15 Soil organic carbon stock in five agriculture land from XGBOOST model

Land use type	Area (ha ⁻¹)	Soil organic carbon stock (ton C ha ⁻¹)				Accuracy Comparison of Validation			
		Mean	Min	Max	SD	RMSE	MAE	R ²	
Rice	2,065,772.48	9.86	0.03	42.49	4.05	4.69	3.49	0.30	
Casava	804,907.12	12.58	0.04	50.16	6.77	6.48	4.31	0.50	
Sugarcane	750,924.29	12.02	0.01	47.78	6.29	6.15	4.37	0.24	
Rubber trees	129,120.46	14.29	0.09	46.67	10.18	6.10	4.10	0.17	
Perennial crop	114,900.75	10.72	0.02	44.41	6.30	6.15	4.37	0.24	
Total	4,218,476.49	-	-	-	-	-	-	-	

When we investigated the maximum value of SOCS on different types of agricultural land, we found that it differed from one location to the next. We found that casava has the highest SOCS, with in at 50.16 ton C ha⁻¹, followed by sugarcane at 47.78 ton C ha⁻¹. Despite the fact that the SOCS values for Rubber trees and perennial crop are nearly identical at 46.67 and 44.41 ton C ha⁻¹, respectively. Rice, on the other hand, has the lowest SOCS of all crops with 42.49 tons of carbon per hectare.

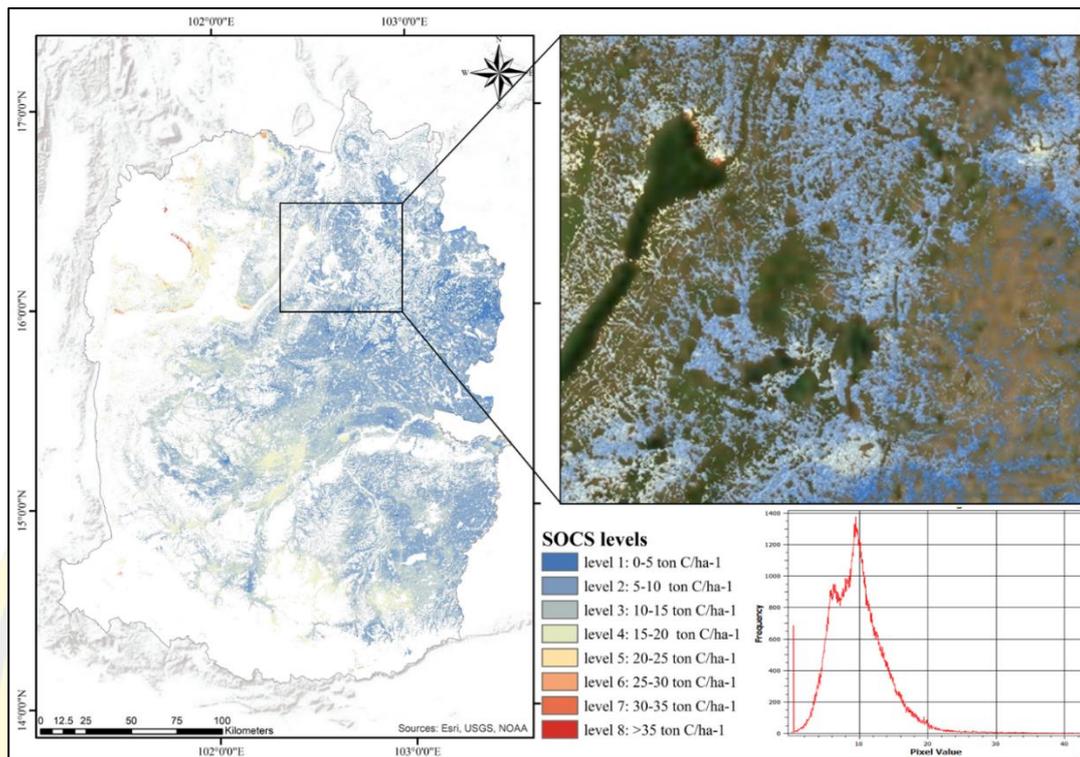


Figure 29 SOCS spatial distribution of paddy field

According to the results, the paddy fields, which comprise approximately 2 million hectares, store the most organic carbon at level 2 (49.71%), with a range of 5-10 tons C ha⁻¹. The majority of these fields are located in the study region's lowlands (Figure 32). The area of level 3 (36.39%) is nearly 751,676.16 ha⁻¹ and is dispersed across the elevated study area. In contrast, level 5-8, which has a high SOCS, has a smaller agricultural field area, show as Table 15.

Table 16 SOCS spatial distribution in difference level of paddy field

SOCS level	Area (ha ⁻¹)	%
Level 1	122,247.73	5.92
Level 2	1,026,801.90	49.71
Level 3	751,676.16	36.39
Level 4	138,079.128	6.68
Level 5	18,523.128	0.90
Level 6	5,916.681	0.29
Level 7	1,948.946	0.09
Level 8	578.803	0.03
Total	2,065,772.48	100

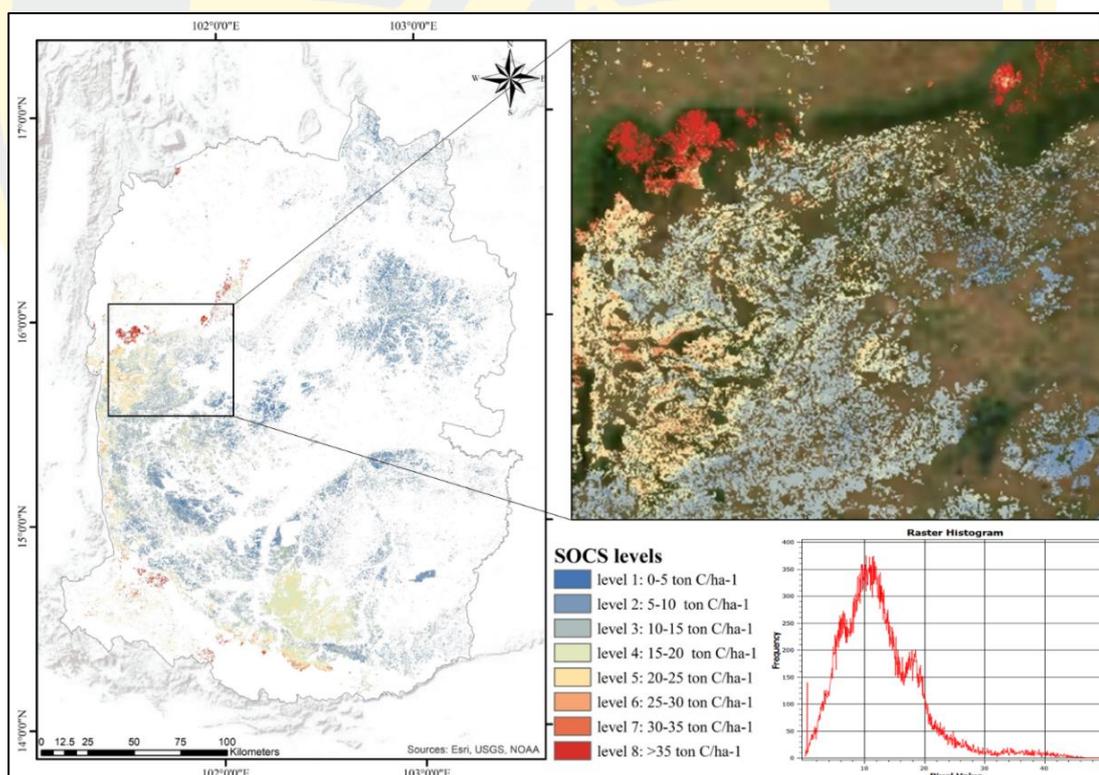


Figure 30 SOCS spatial distribution of casava

The SOCS distribution in cassava fields in the study area is shown in Table 16. The majority of SOCS were found in levels 2-3 (5-15 ton C/ha⁻¹) with 33.77% and 30.52%, respectively (Figure 33), but due to the study area's varied topography, level 2 with 5-10 ton C/ha⁻¹ displays SOCS distribution in the western part of the study area. Additionally, the cassava in the southern part of the studied area contain SOCS between 15 to 20 ton C/ha⁻¹ at level 4.

Table 17 SOCS spatial distribution in difference level of casava

SOCS level	Area (ha ⁻¹)	%
Level 1	63,536.20	7.89
Level 2	245,661.95	30.52
Level 3	271,826.20	33.77
Level 4	162,641.47	20.21
Level 5	34,854.80	4.33
Level 6	8,611.05	1.07
Level 7	6,427.60	0.80
Level 8	11,372.16	1.41
Total	804,907.12	100

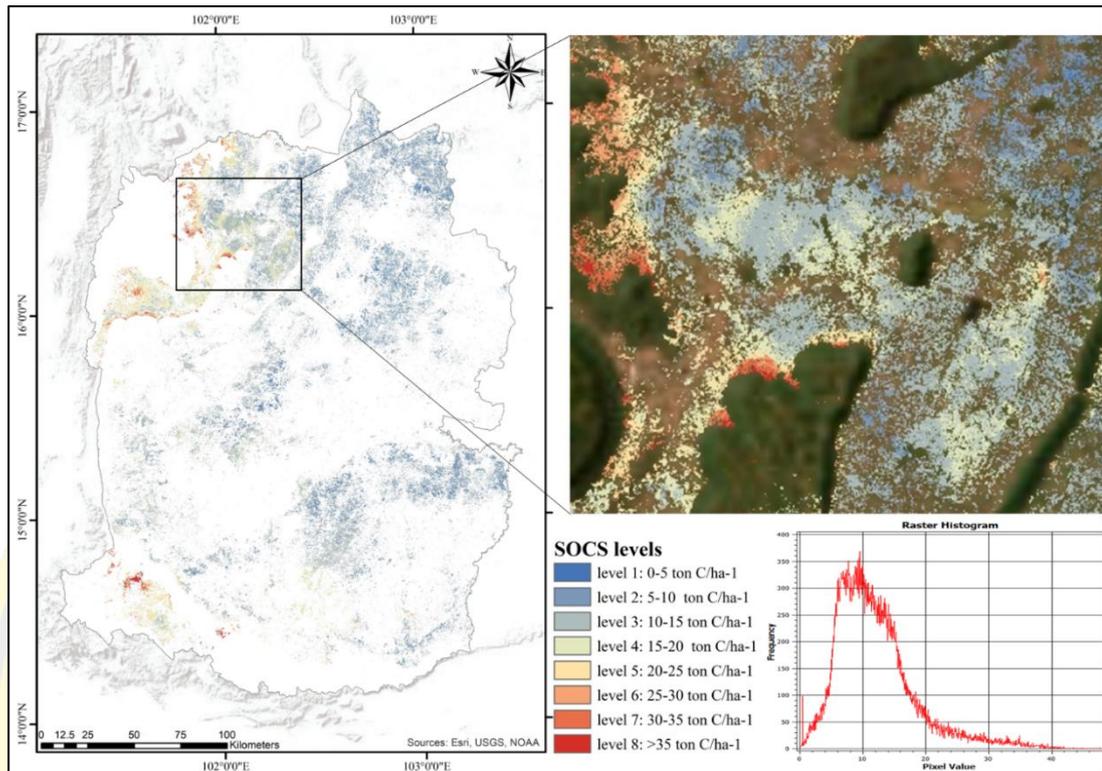


Figure 31 SOCS spatial distribution of sugarcane

Table 17 demonstrates the SOCS in a sugarcane field in difference SOCS level, showing the SOCS distribution with 5-10 tons C/ha^{-1} throughout the entire lowland area and the upland area shows SOCS distribution with 10-20 ton C/ha^{-1} at level 3 and 4, show as Figure 34. In contrast, the lesser area of SOCS distribution in level 8 with >35 tons C/ha^{-1} is located in the western part of the study area near a mountain.

Table 18 SOCS spatial distribution in difference level of sugarcane

SOCS level	Area (ha^{-1})	%
Level 1	45,395.62	6.05
Level 2	277,613.58	36.97
Level 3	260,350.05	34.67
Level 4	96,226.53	12.81

SOCS level	Area (ha ⁻¹)	%
Level 5	38,953.67	5.19
Level 6	21,992.93	2.93
Level 7	6,255.64	0.83
Level 8	4,136.27	0.55
Total	750,924.29	100.00

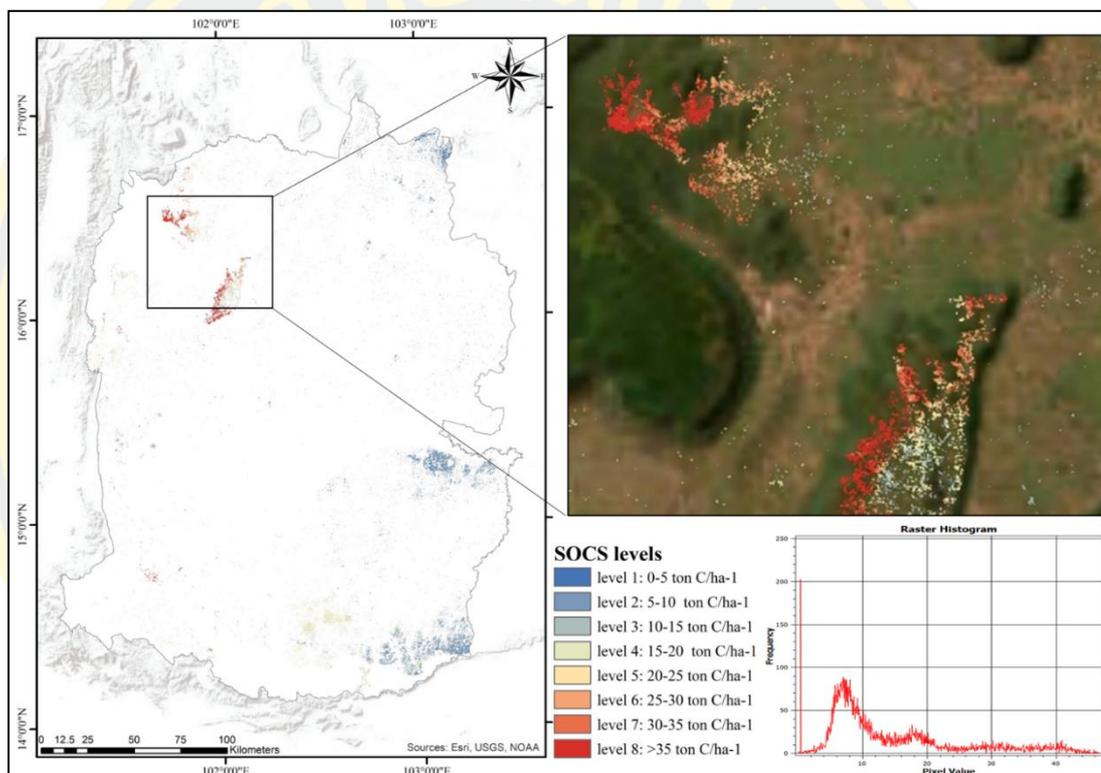


Figure 32 SOCS spatial distribution of rubber tree

According to Table 18, the highest SOCS distribution is level 2 at 59,862.72 ha-1 (5-10 ton C/ha-1), which is clearly visible on the right side of Figure 35 in the north, middle, and south parts. Opposite with level 8, left side on the Figure 35 show SOCS distribution more than 30 ton C/ha⁻¹, covering an area 10,168.41 ha⁻¹. In addition, 10-20 tons C/ha⁻¹ are distributed as levels 3-4 in the southern portion of the study area.

Table 19 SOCS spatial distribution in difference level of rubber

SOCS level	Area (ha ⁻¹)	%
Level 1	6,239.36	4.83
Level 2	59,862.72	46.36
Level 3	20,490.86	15.87
Level 4	18,746.40	14.52
Level 5	4,942.26	3.83
Level 6	4,938.82	3.82
Level 7	3,731.63	2.89
Level 8	10,168.41	7.88
Total	129,120.46	100.00

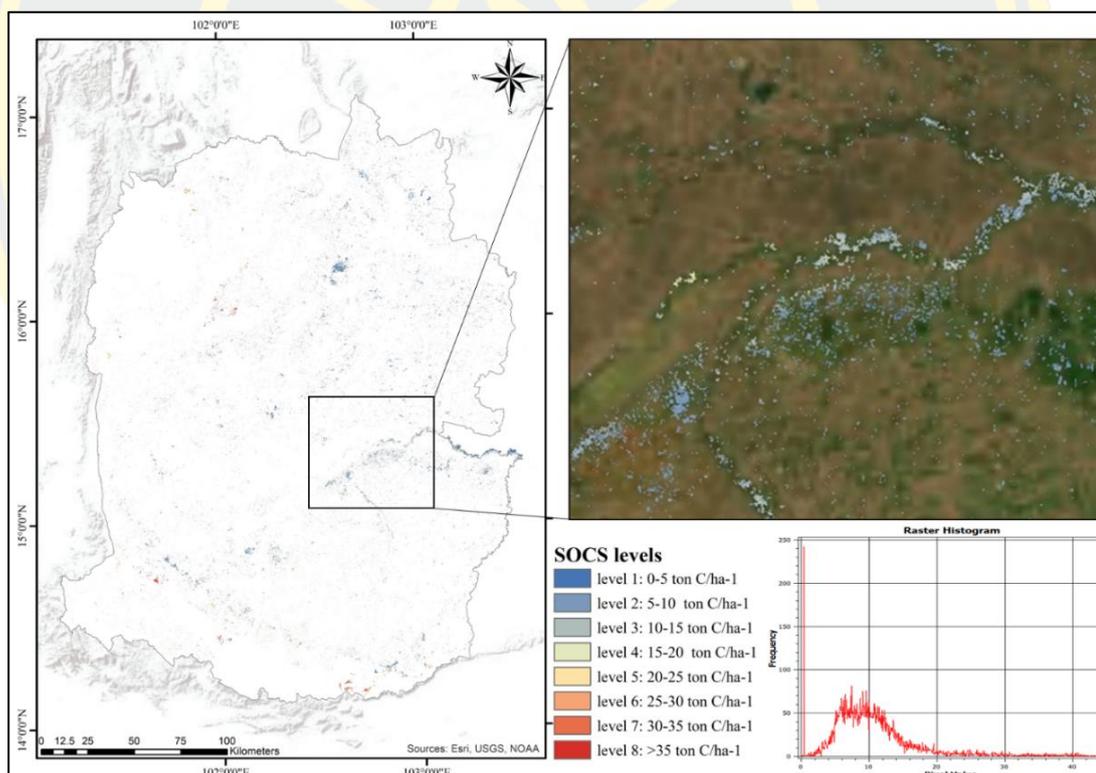


Figure 33 SOCS spatial distribution of perennial crop

Perennial crops, such as Eucalyptus, Teak, Bamboo, Acacia, etc., are distributed in small fields within the entire area under study. Figure 36 shows the

SOCS distribution generally distributed in the area. Level 2 SOCS distributed is the most level that we can see that covered an area of 53,188.29 ha⁻¹, followed by level 3 at 10-15 ton C/ha⁻¹.

Table 20 SOCS spatial distribution in difference level of perennial crop

SOCS level	Area (ha ⁻¹)	%
Level 1	8,274.14	7.16
Level 2	53,188.29	46.01
Level 3	35,405.97	30.63
Level 4	8,921.58	7.72
Level 5	3,796.74	3.28
Level 6	2,199.82	1.90
Level 7	1,502.48	1.30
Level 8	2,319.19	2.01
Total	115,608.21	100.00

In summary, the research areas consist of selected plants such as paddy fields, cassava, sugarcane, rubber tree, and perennial crop on a large agricultural area of 4,218,476.49 hectares. SOCS was estimated using the XGBOOST model with a comprehensive set of environmental variables in various agricultural areas. The result indicates that rice has a lower mean value of SOCS at 9.86 tons C/ha⁻¹; while cassava and sugarcane have mean value of SOCS with approximately 12 tons C/ha⁻¹. Other than that, Rubber trees have the highest mean SOCS value at 14.29 tons C/ha⁻¹. The majority of agricultural land in this study has a SOCS level of 2 to 3, with 5-15 ton C/ha⁻¹, with the exception of cassava, which has the largest amount of land at levels 2 to 4, with 5 to 20 ton C/ha⁻¹.

CHAPTER 5

CONCLUSION AND FUTURE WORK

Conclusion

In this study, the estimation of soil organic carbon stock in the lower region of northeastern Thailand was investigated and analyzed using multisource remote sensing from two sensors (radar and optical) and environmental variables (land use, temperature, precipitation, slope, and elevation) at a resolution of 30 meters after three machine learning algorithms; XGBOOST, RF, and SVR. All the result can conclude according to main three objectives:

1) According to the result of feature importance, slope, vegetation index, soil moisture, and precipitation are significant factors influencing soil organic carbon storage. XGBOOST and RF models demonstrated that slope, EVI, precipitation, soil moisture, and land use are the most predictive factors. Many studies supported this comparable result that SOC content and storage were lower in croplands with steeper slopes. In addition, the SVR model revealed a significant relationship between land use, temperature, BI, and elevation. In contrast, features such as CI, SAVI, SSM, rainfall, VH, red edg4, and EVI mean had negative permutation importance values and did not contribute substantially to the predictions. Several previous studies corroborate the study's findings that the use of Sentinel-1 radar data at C-band in the tropical research region did not yield significant results.

2) Using all remote sensing data and environmental variables, a model was developed to predict soil organic carbon (SOC) levels. The XGBOOST model outperformed previous studies that employed $R^2 = 0.803$, RMSE = 5.569, and MAE = 3.204. Compared to other studies that found XGBOOST to be more accurate than other models, the results were consistent with this finding. According to the results, the XGBOOST model was able to perform the best predictions of SOCS. However, accuracy varied depending on location, size of the study area, population density, and number of soil samples.

3) Utilizing XGBOOST, the soil organic carbon stocks (SOCS) of rice, cassava, sugarcane, rubber tree, and perennial crops were calculated. The carbon

sequestration rates of all plants were at levels 2 and 3, with 10-20 ton C/ha⁻¹, respectively. Mean value of rubber plantations accumulated highest soil organic carbon storage at a rate of 12.52 tons per hectare.

Future work and suggestions

The following are four important recommendations for future research concerning the estimation of soil organic carbon storage using remote sensing data and machine learning algorithms:

Firstly, New remote sensing data from SAR data continue to be fascinating land cover and cloud cover feature variables. Low-frequency SAR data (L-band) is more sensitive to soil parameters like moisture and texture, whereas high-frequency data (X-band) is more sensitive to vegetation attributes. The SMAP L4 carbon product provides an important tool for monitoring and estimating soil carbon levels at regional and global scales.

Secondly, soil samples continue to present the greatest obstacle for regional machine learning model training data. Sample construction of training data must be optimized to reduce the problem of insufficient soil sampling for methodology. Moreover, assessing the land use and land cover change effect on soil organic carbon from landscape to national scale are some outlooks such as climatic zones that can help assess the generalizability for future studies.

Thirdly, new machine learning algorithm and accuracy improvements. Different learning machine algorithms: XGBOOST worked well in this investigation, however other machine learning techniques, such as deep learning, can be tested to predict SOC. Combining data: Remote sensing, soil sampling, and field measurements can improve SOC predictions and understanding of SOC dynamics.

Finally, environmental data variable, incorporating more soil properties: soil texture, incorporating other important soil properties such as soil pH, bulk density, and cation exchange capacity can further improve model accuracy. In addition, integration of crop management practices: Incorporating crop management practices such as tillage, crop rotation, and cover crops into the model can help account for their impact on SOC levels.

REFERENCES

- Ajami M, H. A., Khormali F, Gorji M, Ayoubi Sh. (2016). Environmental factors controlling soil organic carbon storage in loess soils of a subhumid region, northern Iran. *Geoderma*, 1-10.
- Ali, S., Li, Z., Li, X., & Yang, G. (2020). Soil organic carbon stock assessment using spectral indices and machine learning algorithms in a cropland ecosystem. *remote sensing, MDPI*, 12(8)(1272).
- Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. *Remote Sensing*, 11(6), 676. <https://www.mdpi.com/2072-4292/11/6/676>
- annop, P. (2016). *Carbon sequestration of benchmark soil in Northeast Thailand*.
- Annop, P. (2016). *Carbon sequestration of Benchmark Soil in Northeast Thailand*.
- Asner, G. P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environ. Res. Lett.*, 4, 1.
- Azamat SULEYMANOV, I. G., Ruslan SULEYMANOV, Evgeny ABAKUMOV, Vyacheslav POLYAKOV, Peter LIEBELT. (2021). Mapping soil organic carbon under erosion processes using remote sensing. *Hungarian Geographical Bulletin Hungarian Geographical Bulletin* 70, 17.
- Bayartungalag Batsaikhan, O. L., Gantungalag Batsaikhan, Nandinbayar Batsaikhan, Bayanmunkh Norovsuren. (2020). CARBON STOCK ESTIMATION USING REMOTE SENSING DATA AND FIELD MEASUREMENT IN HALOXYLON AMMODENDRON DOMINANT WINTER COLD DESERT REGION OF MONGOLIA. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2020, 1.
- Bilgili, M., Kuzucuoğlu, C., & Özkan, U. Y. (2020). Estimation of soil organic carbon using spectral indices derived from Landsat 8 data: A case study from a semi-arid area of Turkey. *Ecological Indicators*, 110(105912). <https://doi.org/10.1016/j.ecolind.2019.105912>
- Boqiang Xie, J. D., Xiangyu Ge, Xiaohang Li, Lijing Han, Zheng Wang. (2022). Estimation of Soil Organic Carbon Content in the Ebinur Lake Wetland, Xinjiang, China, Based on Multisource Remote Sensing Data and Ensemble Learning Algorithms. *Sensors, MDPI*, 1.
- Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., & van Wesemael, B. (2019). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 267-282.
- Chen, X., Zhang, Y., & Zheng, J. (2019). Responses of soil organic carbon and nitrogen stocks to climate change and land use in a temperate agroforestry system. *Land Use Policy*, 86, 124-131. <https://doi.org/10.1016/j.landusepol.2019.04.033>
- Chih-Chung Chang, C.-J. L. (2022). LIBSVM: A Library for Support Vector Machines. *Department of Computer Science*.
- Demir, G., & Avdan, U. (2021). Assessing the Relationship between Sentinel-1 SAR Data and Soil Organic Carbon Content in Agricultural Lands. *CRC Press*, 13(12)(2387). <https://doi.org/10.3390/rs13122387>
- Dominique Arrouays, N. P. A. S., Hakima Boukir, Claudy Jolivet, Céline Ratié, Marion

- Schrumpf, Lutz Merbold, Bert Gielen, Gaëlle Vincent. (2018). Soil sampling and preparation for monitoring soil carbon. *International agrophysics*, 32, 633-643.
- Donovan, P. (2013). Measuring soil carbon. In *Measuring soil carbon change: A flexible, practical, local method* (pp. 12).
- Duarte, E., Zagal, E., Barrera, J. A., Dube, F., Casco, F., & Hernández, A. J. (2022). Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic. *European Journal of Remote Sensing*, 55:1, 213-231. <https://doi.org/10.1080/22797254.2022.2045226>
- EEA. (2022). Greenhouse gas emissions from land use, land use change and forestry in Europe (8th EAP). Retrieved 11 May 2023, from <https://www.eea.europa.eu/ims/greenhouse-gas-emissions-from-land>
- FAO. (2017). **Soil Organic Carbon the hidden potential**. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS.
- FAO. (2019). *Standard operating procedure for soil total carbon: Dumas dry combustion method*.
- FAO. (2021). *Recarbonizing Global Soils: a technical manual of recommended management practices* 360.
- FAO. (2022). *Global soil organic carbon sequestration potential map (GSOCseq v.1.1) technical report*.
- Fubo Zhao, Y. W., Jinyu Hui, Bellie Sivakumar, Xianyong Meng, Shuguang Liu. (2021). Projected soil organic carbon loss in response to climate warming and soil water content in a loess watershed. *Carbon balance and management, BMC*, 1.
- Gharahi Ghehi, N., Eslamian, S. S., & Moghaddamnia, A. (2019). Soil organic carbon prediction using remote sensing data and machine learning algorithms in a semi-arid region. *Journal of Arid Environments*, 164, 47-57.
- Hartley, I. P., Hill, T. C., Chadburn, S. E., & Hugelius, G. (2021). Temperature effects on carbon storage are controlled by soil stabilisation capacities. *Nature Communications*, 12(1), 6713. <https://doi.org/10.1038/s41467-021-27101-1>
- Huan Wang, X. Z., Wei Wu, Hongbin Liu. (2021). Prediction of Soil Organic Carbon under Different Land Use Types Using Sentinel-1/-2 Data in a Small Watershed. *remote sensing, MDPI*.
- IPCC. (2019). *IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse gas fluxes in Terrestrial Ecosystems*.
- IPCC. (2022). *Climate change 2022: Mitigation of climate change*.
- Jianli Ding, A. Y., Jingzhe Wang, Jingzhe Wang, Danlin Yu. (2018). Machine-learning-based quantitative estimation of soil organic carbon content by VIS/NIR spectroscopy. *PeerJ*. <https://doi.org/10.7717/peerj.5714>
- John, K., Abraham Isong, I., Michael Kebonye, N., Okon Ayito, E., Chapman Agyeman, P., & Marcus Afu, S. (2020). Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land*, 9(12), 487. <https://www.mdpi.com/2073-445X/9/12/487>
- José Padarian, B. M., Alex B. McBratney, Pete Smith. (2021). Additional soil organic

- carbon storage potential in global croplands. *Soil discussion, European geosciences union*, 1.
- K. Nabiollahi, S. E., R. Taghizadeh-Mehrjardi, R. Kerry & J. Triantafyllis. (2019). Assessing soil organic carbon stocks under land use change scenarios using random forest models. *Carbon Management*.
- Klara Dvorakova, P. S., Quentin Limbourg, Bas van Wesemael. (2020). Soil Organic Carbon Mapping from Remote Sensing The Effect of Crop Residues. *remote sensing, MDPI* 1.
- L. A. Jones, J. S. K., N. Madani, R. H. Reichle, J. Glassy, J. Ardizzone. (2016). THE SMAP LEVEL 4 CARBON PRODUCT FOR MONITORING TERRESTRIAL ECOSYSTEM-ATMOSPHERE CO₂ EXCHANGE. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*.
- LDD. (2004a). *Handbook for the analysis of soil samples, water, fertilizers, and plants*. Land development department.
- LDD. (2004b). *The manual of soil water fertilizer and plant analysis for LDD staff*. Land Development Department.
- LDD. (2020). *Land use map*.
- Li, J., Li, Y., Ren, T., & Liang, H. (2016). Estimating soil organic carbon in arid and semi-arid regions using hyperspectral vegetation indices and machine learning algorithms. *Geoderma*, 261 54-66.
- Li, J., Zhou, H., Li, Z., Li, Y., Li, H., Zhao, X., & Chen, Y. (2021). Predicting soil organic carbon content based on environmental variables using support vector regression and artificial neural network. *Catena*, 190(104570).
- Litong Chen, D. F. B. F., Xin Jing, Peter Kühn, Thomas Scholten, Jin Sheng He. (2015). A Comparison of Two Methods for Quantifying Soil Organic Carbon of Alpine Grasslands on the Tibetan Plateau. *PLOS ONE*, 2.
- Manuel Rodríguez-Rastrero, A. O.-M. (2022). Carbon Stock Assessment in Gypsum-Bearing Soils: The Role of Subsurface Soil Horizons. *Earth, MDPI*, 3(839-852), 1.
- Minnikova, T., Mokrikov, G., Kazeev, K., Medvedeva, A., Biryukova, O., Keswani, C., Minkina, T., Sushkova, S., Elgendy, H., & Kolesnikov, S. (2022). Soil Organic Carbon Dynamics in Response to Tillage Practices in the Steppe Zone of Southern Russia. *Processes*, 10(2), 244. <https://www.mdpi.com/2227-9717/10/2/244>
- Montero, D. (2021). eemont: A Python package that extends Google Earth Engine. *Journal of Open Source Software*, 6(62), 3168.
- NSSO. (2018). National Strategy 2018-2037 (Summary). In *National Strategy 2018-2037 (Summary)* (pp. 12). Office of the National Economic and Social Development Board.
- P. Ghimire, B. B., B. Pokhrel, G. Kafle, P. Paudel. (2018). SOIL ORGANIC CARBON STOCKS UNDER DIFFERENT LAND USES IN CHURE REGION OF MAKAWANPUR DISTRICT, NEPAL. *SAARCJ, Agri*, 16(2), 15.
- Pandey, S., Pandey, A. C., Singh, R. P., & Shukla, S. (2017). Remote sensing of soil organic carbon: Prospects and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 58, 14-24. <https://doi.org/10.1016/j.jag.2017.02.008>
- Peng, Q., Guo, L., Huang, Y., Pan, G., Li, L., & Wu, L. (2018). Soil organic carbon

- sequestration in relation to soil properties along a land use gradient in subtropical China. *Catena*, 170, 260-267.
<https://doi.org/10.1016/j.catena.2018.06.015>
- Saini, A. (2021). Support Vector Machine(SVM). *the Data Science Blogathon*.
- Satira Udomsri, J. L. (2015). *Soil resources potential of the northeastern part of Thailand under the project: preliminary development of soil resources database on 2 phased, 1:25,000 orthophotograph*.
- Shi, X., Hu, T., Wu, Y., Xiong, D., & Zhang, Z. (2018). Estimation of soil organic carbon content in a typical karst region using remote sensing and machine learning algorithms. *Remote sensing, MDPI*, 10(4), 574.
- Shuai Wang, M. Z., Qianlai Zhuang, Liping Guo. (2021). Prediction Potential of Remote Sensing-Related Variables in the Topsoil Organic Carbon Density of Liaohekou Coastal Wetlands, Northeast China. *MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.*, 4.
- Sukanlaya Choenkwan, J. M. F., A. Terry Rambo. (2014). Agriculture in the Mountains of Northeastern Thailand: Current Situation and Prospects for Development. *Mountain Research and Development (MRD)*, 97.
- Tao Zhou, Y. G., Cheng Ji, Xiangrui Xu, Hong Wang, Jianjun Pan, Jan Bumberger, Dagmar Haase, Angela Lausch. (2020). Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Science of the Total Environment, Elsevier*, 4-17.
- Theodora Angelopoulou, N. T., Athanasios Balafoutis, George Zalidis , Dionysis Bochtis. (2019). Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. *remote sensing, MDPI*, 11, 676.
- UNFCCC. (1998). *KYOTO PROTOCOL TO THE UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE*.
- V.L. Mulder, S. d. B., M.E. Schaepman, T.R. Mayr. (2010). The use of remote sensing in soil and terrain mapping — A review. *Elsevier*, 162, 12.
- Wang, H., Zhang, X., Wu, W., & Liu, H. (2021). Prediction of Soil Organic Carbon under Different Land Use Types Using Sentinel-1/-2 Data in a Small Watershed. *Remote Sensing*, 13(7), 1229. <https://www.mdpi.com/2072-4292/13/7/1229>
- Wang, J., Yang, J., Liu, X., Li, Y., Zhang, X., & Li, X. (2019). The effects of soil moisture and temperature on soil organic carbon mineralization: A laboratory study. *Journal of Soils and Sediments*, 19(2), 661-669.
<https://doi.org/10.1007/s11368-018-2097-2>
- Were, K., Bui, D. T., Dick, O. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, 52, 394-403.
<https://doi.org/https://doi.org/10.1016/j.ecolind.2014.12.028>
- Xiaohang Li, J. D., Jie Liu, Xiangyu Ge, Junyong Zhang. (2021). Digital Mapping of Soil Organic Carbon Using Sentinel Series Data: A Case Study of the Ebinur Lake Watershed in Xinjiang. *Remote Sens, MDPI*, 19.
- Xie, B., Ding, J., Ge, X., Li, X., Han, L., & Wang, Z. (2022). Estimation of Soil Organic Carbon Content in the Ebinur Lake Wetland, Xinjiang, China, Based on

- Multisource Remote Sensing Data and Ensemble Learning Algorithms. *Sensors*, 22(7), 2685. <https://www.mdpi.com/1424-8220/22/7/2685>
- Yangchengsi Zhang, L. G., Yiyun Chen, Tiezhu Shi, Mei Luo, QingLan Ju, Haitao Zhang, Shanqin Wang. (2019). Prediction of Soil Organic Carbon based on Landsat 8 Monthly NDVI Data for the Jiangnan Plain in Hubei Province, China. *MDPI, Remote Sens*, 21.
- Yin Gao, Lijuan Cui, Bing Lei, Yanfang Zhai, Tiezhu Shi, Junjie Wang, Yiyun Chen, Hui He, & Wu, G. (2014). Estimating soil organic carbon content with visible-near-infrared (vis-NIR) spectroscopy. . *Appl Spectrosc*, 68(7)(712), 22. <https://doi.org/10.1366/13-07031>
- Zhang, X., Liu, G., Yan, J., Wang, J., & Zhang, X. (2019). Slope gradient and aspect control soil organic carbon storage and quality in cropland on the Loess Plateau of China. *Journal of Arid Land*, 11(2), 251-263.
- Zhang, X., Liu, Y., Shao, M., & Zhao, Y. (2017). Estimation of soil organic carbon stock using remote sensing data and a machine learning algorithm in a dryland region of China. . *Geoderma*, 289.
- Zhao, X., Shen, Y., Wu, J., & Zhang, X. (2017). Soil carbon sequestration under different agricultural practices: A review. *Journal of Integrative Agriculture*, 16(11), 2558-2570. [https://doi.org/10.1016/S2095-3119\(17\)61797-0](https://doi.org/10.1016/S2095-3119(17)61797-0)
- Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., & Lausch, A. (2021). Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Science of The Total Environment*, 755, 142661. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142661>
- Zomer, R. J., Bossio, D. A., Sommer, R., & Verchot, L. V. (2017). Global Sequestration Potential of Increased Organic Carbon in Cropland Soils. *Scientific Reports*, 7(1), 15554. <https://doi.org/10.1038/s41598-017-15794-8>

BIOGRAPHY

NAME Ms. Sunantha Ousaha

DATE OF BIRTH March 5, 1991

PLACE OF BIRTH Banpong, Ratchaburi Province, Thailand

PRESENT ADDRESS 174 Vilege no.5, Krab Yai Sub-district, Banpong District, Ratchaburi Province, Thailand

POSITION HELD 2015-2016 Expert assistance at Land Development Regional Office 1, Land Development Department
2016-2017 Agricultural research officer at soil resources survey and research division, Land Development Department
2017-present Soil surveyor, practitioner level at soil resources survey and research division, Land Development Department

EDUCATION 2009-2013 Bachelor's degree in soil science(B.S.), Faculty of Agriculture at Kamphaeng Sean, Kasetsart University, Thailand

AWARDS OR GRANTS 2012 Certificate of academic excellence, department of soil science(B.S.), Kasetsart University, Thailand
2021 Ministry of Science and Technology Scholarship for Double Master Degree on Geo-Informatics (SCGI Master Degree Program)